



सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

21805
217

I.A.R.I.6.

GIP NLK—H 3 I.A.R.I.—10-5 55—15,000

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 188

LONDON

Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W.1

25 February 1947

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
(Cambridge, and Bentley House, London)
Agents for Canada and India: Macmillan

CONTENTS

SERIES A VOLUME 188

No. A 1012. 31 December 1946

	PAGE
Bakerian Lecture. The more recent work on the reaction between hydrogen and oxygen. By C. N. Hinshelwood, F.R.S.	1
A general kinetic theory of liquids. I. The molecular distribution functions. By M. Born, F.R.S. and H. S. Green	10
Some applications of the Riesz potential to the theory of the electromagnetic field and the meson field. By N. E. Fremberg	18
The crystal structure of some molecular complexes of 4:4'-dinitrodiphenyl. I. The complex with 4-hydroxydiphenyl. By D. H. Saunderson	31
The crystal structure of meta-dinitrobenzene. By E. M. Archer	51
Measurement of the ultimate pressures of oil-diffusion pumps. By J. Bleas	62
The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. By A. D. Booth	77
The mechanism of the catalytic oxidation of ethylene. I. Experiments using a flow system. By G. H. Twigg	92
The mechanism of the catalytic oxidation of ethylene. II. Reactions between ethylene, etc. and chemisorbed oxygen monolayers. By G. H. Twigg	105
The mechanism of the catalytic oxidation of ethylene. III. The reactions between ethylene and oxygen in a static system. By G. H. Twigg	123
Corrigendum	141

No. A 1013. 30 January 1947

Anniversary Address by Sir Robert Robinson	143
The theory of the Raman effect in crystals, in particular rock-salt. By M. Born, F.R.S. and M. Bradburn	161
Thermal scattering of X-rays by crystals. I. Dynamical foundation. By G. H. Begbie and M. Born, F.R.S.	179
Thermal scattering of X-rays by crystals. II. The thermal scattering of the face-centred cubic and the close-packed hexagonal lattices. By G. H. Begbie	189
The aerodynamic drag of the earth's surface and the value of von Karman's constant in the lower atmosphere. By P. A. Sheppard. (Plates 1 and 2)	208
The structure of polyisoprenes. VI. An investigation of the molecular structure of dibenzyl by X-ray analysis. By G. A. Jeffrey	222
The internal ballistics of a leaking gun. By J. Corner. (Plate 3)	237
The internal ballistics of a gun after shot ejection. By J. Corner	255
The theory of wedge indentation of ductile materials. By R. Hill, E. H. Lee and S. J. Tupper. (Plate 4)	273

No. A 1014. 11 February 1947

PAGE

The detonation of liquid explosives by gentle impact. The effect of minute gas spaces. By F. P. Bowden, M. F. R. Mulcahy, R. G. Vines and A. Yoffe	291
The period of impact, the time of initiation and the rate of growth of the explosion of nitroglycerine. By F. P. Bowden, M. F. R. Mulcahy, R. G. Vines and A. Yoffe. (Plates 5 to 9)	311
Hot spots on rubbing surfaces and the detonation of explosives by friction. By F. P. Bowden, M. A. Stone and G. K. Tudor. (Plates 10 and 11)	329
The quantal calculation of the photo-ionization cross-section of atomic potassium. By D. R. Bates	350
Structure and thermal properties associated with some hydrogen bonds in crystals. VII. Behaviour of KH_2PO_4 and KH_2AsO_4 on cooling. By A. R. Ubbelohde and I. Woodward. (Plate 12)	358
The mechanism of corrosion fatigue of mild steel. By U. R. Evans and M. Tchorabdj Simnad	372
The electric conductivity and the activation energy of ionic migration of molten salts and their mixtures. By H. Bloom and E. Heymann	392
An electron diffraction study of the structure of thin films of normal paraffins. By K. G. Brummage. (Plate 13 and 14)	414

No. A 1015. 25 February 1947

The anisotropy of thermal expansion as a cause of deformation in metals and alloys. By W. Boas and R. W. K. Honeycombe. (Plates 15 to 18)	427
The two-dimensional hydrodynamical theory of moving aerofoils. IV. By R. M. Morris	439
Transverse magnetization in ferromagnetic crystals in relation to domain structure. By A. von Engel and M. S. Wills	464
Electron traps and dielectric changes in phosphorescent solids. By G. F. J. Garlick and A. F. Gibson	485
The effect of temperature on the intensity of X-ray reflexion. By E. A. Owen and R. W. Williams	509
On the theory of dielectric breakdown in solids. By H. Fröhlich.	521
Energy distribution and stability of electrons in electric fields. By H. Fröhlich	532
Intensities of γ rays, studied by means of their Compton secondaries. By K. Siegbahn	541
Index	549

BAKERIAN LECTURE

The more recent work on the reaction between hydrogen and oxygen†

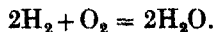
BY C. N. HINSELWOOD, F.R.S.

(Delivered 11 July 1946—Received 22 July 1946)

It is not infrequently in the evolution of a scientific problem that the following course of events is observed. First, there is a preconceived notion about the way in which a certain phenomenon should occur. Next, this idea is rather rudely dispelled by experiments which reveal all sorts of unsuspected complexities. The subject seems to become more and more difficult and stimulates a good deal of effort and contributions from many sides. At this stage the clearest way of treating the matter is usually to approach it historically, or at any rate analytically, and even then expositions of it usually give the impression of being accessible only to specialists. Gradually, however, things clarify, the complexities seem in an increasing degree to assume the guise of details which can be derived as consequences from the general theory, and a synthetic treatment becomes possible. In the light of all the intervening work it almost appears as though everything could, from the start, have been deduced from first principles. Although this appearance may from one point of view be illusory, it is none the less a sign that the task is approaching completion.

In the development of modern ideas on chemical kinetics the study of gaseous reactions has played an interesting part. Many unexpected and sometimes disconcerting observations have shown that the kind of relations which might have been assumed in the light of the earlier ideas do not exist, and yet one begins to see that the tangle of facts has after all a coherence of its own, though quite different from what was first imagined.

Among gas reactions the combination of hydrogen and oxygen to form steam has proved not the least fascinating. The chemical equation is written in the form



According to preconceived ideas based upon over-hasty application of the law of mass action, this reaction was expected to occur at a rate proportional to the square of the concentration of hydrogen and to the first power of the concentration of oxygen.

† During the current year various papers on this subject by Dr A. H. Willbourn, Mr C. F. Cullis and myself have appeared in the *Proceedings*. This lecture serves in part as a commentary on the whole series, and I should like to take this opportunity of expressing very deep indebtedness to the colleagues whose work is discussed. References to the work of other authors, including especially the recent important publications of Lewis and von Elbe, are given in the individual papers mentioned above, viz. *Proc. Roy. Soc. A*, 1946, 185, 353, 369, 376; 186, 462, 469. (In Part I the constant in equation (16) should be 3.85×10^{-7} instead of 2.64×10^{-6} , a clerical error which does not affect subsequent calculations.)

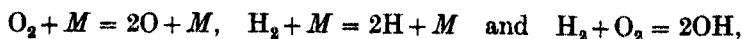
An historical account and discussion of the various researches—English, Russian and American—which dispelled this idea and revealed, or nearly revealed, the true mechanism would be rather lengthy, and would certainly be complicated to follow except for those with a specialized interest. In recent years, partly under the stimulus of practical problems connected with the ignition of gases from propellants, this classical reaction has been the subject of renewed study, and a good deal of fresh evidence has become available. It seems expedient, therefore, to give a summarizing account and in particular to attempt the task of presenting the matter synthetically, so that the complexities fall into their true perspective as details, and the essential simplicity of the mechanism is allowed to appear.

If we approach the problem of the hydrogen-oxygen reaction in the light of modern ideas we realize from the start that it must almost certainly prove to be a chain reaction. This is because both hydrogen and oxygen are capable of resolution into atoms by reactions with large but by no means unattainable activation energies, and because free atoms or radicals once formed are known to require very small activation energies to attack other molecules. The initiation of the chain reaction requires a high activation energy and is unlikely: but chains once started have a very good chance of being propagated to a great length. For any chemical reaction there is in principle a competition between a purely molecular process which is easy to start but does not propagate itself, and a free radical chain process which is difficult to start and propagates itself very effectively. Often enough the two kinds of mechanism make contributions of the same order of magnitude to the observed reaction, just as evolutionary and revolutionary methods in the political sphere may lead in the end to much the same results. But in a highly exothermic reaction the balance is likely to prove decisively in favour of the chain mechanism. We have outlived the gratification of the original discovery that the hydrogen-oxygen reaction constitutes a beautiful example of such a mechanism. We should nowadays feel able to take it for granted almost without investigation.

Various general principles now come to our aid. In the first place the modern knowledge of bond energies places us in a favourable position for estimating the likelihood of any given reaction in which an atom or free radical participates or is formed.

Another principle which has become rather definitely established is that reactions nearly always resolve themselves into a series of what amount to the simplest possible steps, each step involving the minimum disturbance of existing bonds and structures. Examples of this have accumulated from many sides—the pyrolysis of hydrocarbons, the actions of acids on metals, and so on—though it is doubtful whether the full implications of the principle have yet been exploited.

In the light of this idea, a mechanism corresponding to the chemical equation and involving two molecules of hydrogen and one of oxygen hardly calls for consideration. For the initiation of the chain there is an overwhelming likelihood of a bimolecular collision process. Of these there are only three to choose from:



where M represents any molecule present.

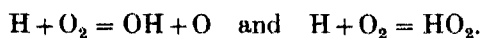
The first is very unlikely compared with the second, since it would require considerably more energy. The choice between the remaining two will be deferred for the moment, first because it constitutes one of the subtler problems which can only be dealt with when the rest of the mechanism is clear and, secondly, because for most of the phenomena shown by the hydrogen-oxygen system it is immaterial.

Now let us suppose that hydrogen atoms and hydroxyl radicals are brought into a mixture of hydrogen and oxygen, and consider what their probable fates will be.

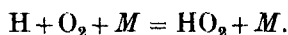
Hydroxyl radicals are only likely to react with hydrogen molecules and hydrogen atoms with oxygen molecules. There is not much for the former to give except



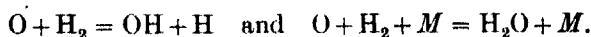
With the latter there are two possibilities:



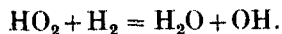
Known principles of physics now tell us that a bimolecular association reaction is very unlikely with such simple particles as H and O₂, which, if they combine, require the presence of a third particle to remove the energy of combination:



In so far as ternary collisions may occur in the system, this reaction is possible. The oxygen atoms produced in the first of the alternatives are only likely to react with hydrogen molecules, and again there are alternative possibilities, corresponding to those for the hydrogen atoms:

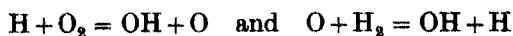


No other atoms or radicals seem likely to be formed. Further, the reactions of the hydrogen atoms, oxygen atoms and hydroxyl radicals can be repeated indefinitely so that a chain process is seen to be practically certain. The only further change, apart from recombination reactions, likely to be suffered by any of the species present is



Something can also be predicted about the termination of the chains. The most obvious way is by the recombination of two atoms or radicals. As a matter of fact, however, such reactions are at a great disadvantage compared with possible competitors, because they necessitate not only the encounter between two species both present in extremely minute concentration but, in addition, the presence of a third particle. If reaction chains can be ended in alternative ways involving the participation of only one free atom or radical, these ways will in general have an easy preference. Such a way does exist in the diffusion of the active radicals or atoms to the vessel wall where they are absorbed and removed from further participation in the reaction cycles. And in fact, as might be expected, all the processes which finally terminate chains in this system prove to depend upon such loss by diffusion of H, OH, O or HO₂.

The most characteristic of the phenomena shown by the hydrogen-oxygen reaction, however, depend not upon the mere existence of the chains but upon the fact that they branch. Each of the reactions

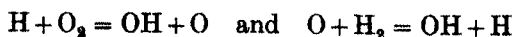


gives two radicals for each one initially entering. The speed of the combination as a whole must therefore build up with immeasurable rapidity to explosion, except in so far as the branching of the chains can be held in check. Everything, therefore, depends upon the balance between the branching of the chains and their control by chain-breaking processes.

Since the total production of new radicals depends upon the volume, while the total loss by diffusion is proportional to the surface of the vessel, one must expect the system to show in marked degree what is known as a scale effect. As the size of the vessel increases, the volume effect increases more rapidly than the surface effect, and if it is initially smaller, there must be a perfectly well-defined critical limit of size at which the branching effect just surpasses the diffusion effect. At this point a controllable reaction would be superseded by an explosion. This phenomenon of the critical limit is indeed found to be very much in evidence. Abrupt transitions from slow reaction to explosion began nearly two decades ago to interest chemists who made many detailed studies of the remarkable phenomena depending upon the scale effect in systems with branching chains. It is probably fair to say that their investigations quickened the appreciation of the possibilities in the even more spectacular example which has recently occupied the attention of physicists.

The scale effect will appear not only in relation to vessel size, but also in relation to pressure. Diffusion of atoms or radicals to the walls of the vessel decreases as the pressure of gas increases. One might therefore, at a given temperature, expect a controllable reaction rate at the lowest pressures succeeded by inflammation and explosion above a critical pressure limit. This indeed is observed and at 550° C, in a vessel of a few hundred c.c. capacity, the so-called first explosion limit is observable at pressures of the order of 1 mm. The variation of the critical pressure depends upon the partial pressure of any gases present and upon the size of the vessel in the manner predictable from the influence of these factors on the loss of chain-propagating particles by diffusion to the wall. Numerous studies of the transition from quiet reaction to explosion have been made not only for the hydrogen-oxygen system, but for systems of the most varied kind, and on the whole have given excellent confirmation of theoretical ideas.

This simple picture of a region of quiet reaction followed by one of explosion might have been complete but for a factor inherent in the chemistry of the system. The two chain-branching processes



are in competition with the association processes



Both the latter become relatively more important the higher the pressure. Both quench the branching of chains, so that with increasing pressure, the so-called second explosion limit is reached, where explosion gives place on further increase of pressure to controllable reaction. At 550° C this occurs at approximately 100 mm. in a mixture of $2\text{H}_2 + \text{O}_2$.

This second limit formed the subject of much investigation which fully confirmed that the controlling process occurred in the gas phase. Since the rate of branching of chains is proportional to a binary collision product, $[X][Y]$, where $X = \text{H}$ or O and $Y = \text{O}_2$ or H_2 , and since the rate of quenching is proportional to the ternary collision product of the same X and Y with any other molecule, we have as a condition for the limit

$$K[X][Y] = K'[X][Y]\{a[\text{H}_2] + b[\text{O}_2] + c[M]\},$$

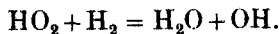
whence

$$[\text{H}_2] + k_{\text{O}_2}[\text{O}_2] + k_M[M] = K^*. \quad (1)$$

This is a very characteristic form and is fully confirmed by experiment. The relative values of the constants, k_{O_2} and k_M , can be calculated with some measure of success from the kinetic theory, on the assumption that, for molecules of the same atomicity, they correspond to the relative frequencies of ternary collisions. From the temperature coefficient of the second explosion limit the activation energy of the branching process follows immediately. It comes out to be 26,000 calories.

The knowledge of this quantity enables us to see something rather important about the whole scale of the phenomenon. With this activation energy at 550° C one collision in about 10^7 should lead to branching. The kinetic theory tells us that in a vessel of 5 cm. radius about 1.5×10^{10} collisions could be made by particles diffusing at atmospheric pressure from the middle of the vessel to the wall. Branching could occur many times and explosion would be certain unless the ternary collision quenching controlled it. Control by diffusion to the wall becomes predominant when the number of collisions made by a particle on its way there is of the order 10^7 . This occurs at the pressure of $10^7/1.5 \times 10^{10}$ atm., or about 0.5 mm. This is in fact just the order of magnitude of the first pressure limit, so that our 'predictions' are now fulfilled.

Returning to the question of control of branching, the two association reactions differ in that one of them gives the inert molecule H_2O , and thus finally ends any possible radical chain, while the other gives HO_2 , which must itself be a very reactive particle. The only molecule with which it is likely to react, however, is hydrogen, so that we can predict with some assurance the following transformation:



Thus the formation of HO_2 in a ternary collision, although it involves the quenching of branching, does not terminate a chain. Indeed, the final result of the quenching is only that H atoms are replaced by an equivalent number of OH radicals. HO_2 can and does diffuse to the wall, where it is destroyed. In so far as this happens, the quenching process is final; but, the higher the total pressure in the system, the smaller does this diffusion become.

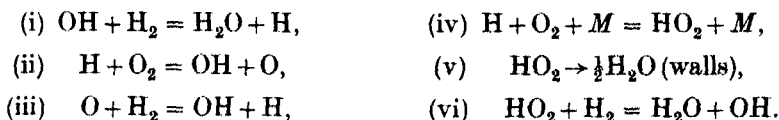
As the destruction of HO_2 in this way becomes less important and its reaction with hydrogen more important, we arrive at a state of affairs where the HO_2 formation

merely delays without finally inhibiting the branching of the chains. At high enough pressures, therefore, a third explosion limit is to be expected. The region of pressure between the second and third limits represents a mere interlude, caused by the intervention of the ternary collision processes, in what would otherwise have been a simple progress from non-explosion to explosion. During this interlude, however, there occur extremely interesting phenomena to which reference will presently be made.

The third explosion limit is in fact observed experimentally, but for a long time its nature remained in some doubt. The difficulty is that at pressures just below the limit, the reaction velocity is extremely high. Even if there were no branching chains representable by chemical equations, there could easily be serious departures from thermal equilibrium, and the explosion might result from a mere breakdown of isothermal conditions. The most recent work, however, leaves little doubt that the third limit depends upon the chain branching in the way expected. This matter has been discussed in some detail in very recent papers in the *Proceedings* and the results may be summarized by saying that the influence on the limit of (a) hydrogen-oxygen proportions, (b) addition of nitrogen, carbon dioxide and steam, are all reasonably consistent with the theory that an essentially isothermal branching of chains occurs. The various constants involved are in fairly satisfactory agreement with what might be expected from the kinetic theory.

We may now collect together the chemical equations expressing the processes which we have supposed to occur, namely:

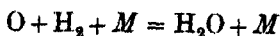
initiating reaction giving H or OH (rate = f_1):



This leads, by the usual methods, to the following expression for the rate of formation of water in the steady state:

$$\frac{d[\text{H}_2\text{O}]}{dt} = \frac{f_1 \left\{ \frac{1.5k_5 + 2k_6[\text{H}_2]}{k_5 + k_6[\text{H}_2]} \right\}}{1 - 2k_2/\Sigma k_4[M] - k_6[\text{H}_2]/(k_5 + k_6[\text{H}_2])}. \quad (2)$$

In the series of chemical equations from which (2) is derived, certain omissions have been made for simplicity. First, the diffusion to the wall of all chain carriers except HO_2 has been neglected. This is justified by the fact that the first limit occurs at very low pressures compared with the second and third. The simplification means, however, that equation (2) describes the second and third limits but does not pretend to include the first, which is best treated quite separately. The separate treatment is, fortunately, quite possible, since in the neighbourhood of the first limit the processes determining the second and third are negligible. The next simplification is that of the two possible ternary quenching processes only (iv) has been written down and is taken as representative of both. This is reasonable since the alternative



would have no aftermath corresponding to (v) and (vi). In any event the procedure seems to be justified by the results.

The condition for explosion is that the denominator in equation (2) should equal zero. For given proportions of H_2 , O_2 and M , this condition is satisfied in general by two values of the pressure, one corresponding to the second limit and the other to the third. The constants k_4 and k_5 are calculable respectively from k_{O_2} and k_M of equation (1) and from the relative diffusion coefficients of HO_2 . The latter may be expressed in an inverse form as

$$D'_M = D_{H_2}/D_M.$$

The following comparison then emerges:

M	D'_M		k_M		
	experimental from equation (2)	from kinetic theory	experimental from equation (2)	experimental from equation (1)	from kinetic theory
H_2	1.0	1.0	1.0	1.0	1.0
O_2	3.0	3.8	0.38	0.36	0.40
N_2	2.2	3.9	0.30	0.37	0.45
CO_2	2.5	5.2	0.90	0.90	0.51
H_2O	4.0	4.1	8.3	9.1	0.62

In commenting on the above table one may say the following: the values of D'_M of the heavier gases relative to hydrogen are a significant confirmation of theory. The importance of the deviations among the heavier gases themselves is difficult to assess. The agreement of the values of k_M obtained from the equations (1) and (2) respectively is very important and shows that the third limit and the second limit are governed by the same mechanism. This seems to establish the explosion at the third limit as a phenomenon depending on branching chains. The relative values of k_M for hydrogen, oxygen and nitrogen can be more or less correctly calculated from the kinetic theory: the same applies to the relative values for helium and argon. Comparison, however, between gases of different atomicities does not give satisfactory results. The reason is that, for the simple gases of the same atomicity, the relative quenching is proportional to the relative numbers of ternary collisions made, while for gases of different atomicity, or for more complex molecules, the quenching power depends upon specific interactions.

The general form of (2) is in very satisfactory agreement with experiments on the third limit. The conditions for explosion derived from equation (2) are, however, quite independent of the form of the function f_1 , which describes the mode of initiation of the chains. For this very reason the latter has proved somewhat difficult to determine and is perhaps still the matter on which most differences of opinion would exist. The view which will be expressed here, though given with due reserve, is based upon what would seem to be the most direct approach. It was mentioned above that between the second and third explosion limits there occurs an interlude of measurable reaction. During this, interesting phenomena are observed, the reaction rate depending in the most remarkable ways on the concentrations of all the gases present, as well as upon the dimensions of the vessel. The effects for a long time proved difficult to analyse.

But a great simplification has occurred since the problem of the third limit has been more or less cleared up. The study of the latter gives us all the terms in the denominator of equation (2), together with all the terms in the numerator except f_1 . By combining, therefore, measurements on the third limit with measurements of the reaction velocity between the second and third limits, we can test directly the form of the function f_1 . The best fit seems to be obtained by writing

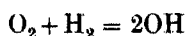
$$f_1 = K[\text{H}_2] \cdot \{Z_{\text{H}_2}[\text{H}_2] + Z_{\text{O}_2}[\text{O}_2] + Z_M[M]\},$$

where Z_{H_2} , Z_{O_2} and Z_M are the relative collision numbers for hydrogen molecules respectively with hydrogen, oxygen and any inert gas M . This expression is based upon the idea that chains are started by the dissociation of hydrogen molecules in collision with any other molecule present.

It must be emphasized that this result is found under certain definite conditions, namely between 570 and 600° C, and in a vessel whose surface is coated with potassium chloride. At lower temperature there seems little doubt that the chains have their origin largely on the wall of the vessel.

Once again by combining measurements on the third limit with determination of reaction rate the complex knot of temperature effects can be unravelled in a relatively simple way, though the results are only approximate since the range of temperature over which one can work is very limited. Under the conditions mentioned, the detailed study of temperature effects gives for the energy of activation of the chain-initiating process a value of 90–110 kcal., which is fully consistent with the hypothesis of dissociation of hydrogen molecules with atoms, the energy required for this process being 101 kcal.

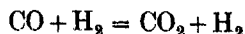
It would not be right to convey the impression that no problems are now left about the mechanism of the hydrogen-oxygen combination. Some authors have given reasons for preferring



as the initial step: there are certain complexities about the influence of steam on the reaction, and also problems about the development with time of the reaction in its initial stages. But it is probably fair to say that all these matters when resolved would leave the general picture not very different from what has been sketched.

It will now be convenient to consider briefly some other aspects of this remarkable reaction.

One fact to which reference might be made is that the oxidation of carbon monoxide in presence of small amounts of moisture has long been known to proceed by the indirect mechanism:

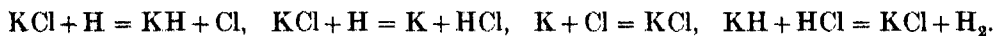


followed by the reoxidation of the hydrogen. The whole reaction, as a result, shows several characteristics, including a marked susceptibility to inhibition by halogens, which it derives from the hydrogen-oxygen reaction.

The action of alkali salts in suppressing the 'flash' given by propellants has long been known and almost certainly depends upon the quenching of hydrogen ignition.

It was soon discovered that the speed of combination of hydrogen and oxygen may itself be reduced a hundred times or more by coating the reaction vessel with potassium chloride. Many salts of alkalis and alkaline earths have the same effect. The first and third explosion limits, which depend upon the surface, are profoundly affected by the salts: the second limit which is controlled by the gas-phase process is not. There is little doubt, therefore, that the action depends upon the efficient removal of radicals at the surface, and the question arises as to its mechanism.

One of the principles which detailed investigation seems to be establishing more and more clearly is that catalytic or anticatalytic mechanisms are usually representable by definite chemical equations, and, moreover, by chemical equations resolvable into a series of steps of the simplest possible kind. Hydrogen atoms are known to be among the more important participants in the hydrogen-oxygen reaction. Most, if not all, of the metals whose salts are active form hydrides. It seems very likely that definite cycles of reaction occur, such as the following:



This can lead to a catalytic recombination of hydrogen atoms without permanent change of the salt. A result which at first sight seems rather peculiar is found if the walls of the vessel are coated with potassium iodide. The mechanism undergoes a complete transformation. The chain reaction is replaced by a surface reaction, and the pronounced dependence of the rate upon the pressures of hydrogen, oxygen and inert gas, observed in chloride-coated vessels, disappears. The rate is now *independent* of the hydrogen pressure, independent of added nitrogen, and shows a dependence on oxygen pressure of the form $a + b[\text{O}_2]$. What appears to happen is that in a cycle of changes similar to that just written down for potassium chloride, free iodine escapes from the surface in minute quantity. Iodine, at pressures of a fraction of a millimetre has long been known to cause complete suppression of the chains in the gas phase, which it does by removing hydrogen atoms.

Even if chlorine escaped from the chloride-coated surface in a similar way it would not interfere seriously with the chains, its effect being known to be almost negligible compared with that of iodine, and for understandable reasons. The somewhat surprising result found in the potassium iodide-coated reaction vessel thus provides indirect confirmation of the idea that the halide salts enter with the gases into definite cycles of chemical change.

It must not be suggested that the last word has by any means been said about this wonderful reaction between two of the commonest gases. There are, as was remarked above, many matters of detail about which doubts and differences of opinion exist. Nevertheless it can be said that the general framework is established. In this lecture I have discussed it from the standpoint of first principles and tried to show how most of the results could be deduced from them. Actually, of course, the study of the hydrogen-oxygen reaction has played no small part in the understanding of those principles themselves.

A general kinetic theory of liquids

I. The molecular distribution functions

BY M. BORN, F.R.S. AND H. S. GREEN, *Edinburgh University*

(Received 11 February 1946)

This paper outlines a general theory whose object is to provide a basis from which all the equilibrium and dynamical properties of liquids can be investigated.

A set of multiform distribution functions is defined, and the generalized continuity equations satisfied by these functions are derived. By introducing the equations of motion, a set of relations is obtained from which the distribution functions may be determined. It is shown that Boltzmann's equation in the kinetic theory of gases follows as a particular case, and that, in equilibrium conditions, the theory gives results consistent with statistical mechanics. An integral equation for the radial distribution function is obtained which is the natural generalization of one obtained by Kirkwood for 'rigid spherical molecules'. Finally, it is indicated how the theory may be applied to solve both equilibrium and dynamical problems of the liquid state.

1. INTRODUCTION

It has been said that there exists no general theory of liquids because it is impossible to utilize the simplifying conditions either of the kinetic theory of gases (Chapman & Cowling 1939) where the density is small, or of the theory of solids (Born 1923) where a high degree of spatial order may be assumed. Yet a mathematical formulation of the problem should be possible, without making such an assumption, since only the general laws of mechanics and statistics are involved; though the solution itself may be extremely difficult.

In the following we shall give a set of equations which describe not only the equilibrium of a statistical assembly of molecules—this could be done equally well by the methods of statistical mechanics—but also the dynamical properties. Our method consists of the introduction of not a single distribution function, as in the kinetic theory of gases, but a complete set of such functions for singlets, doublets, triplets, etc., of molecules in the assembly, and depending not only on position and velocity, but also on higher derivatives with respect to the time.

The latter are eliminated with the help of the equations of motion. The only previous publications known to us where this device is employed are due to A. Vlasov (1945), but his work can be regarded only as in the right direction, as it is rendered unconvincing by the introduction of certain arbitrary approximations at an early stage. The idea of multiple distribution functions is also not entirely new; for example, the investigation of the scattering of X-rays by liquids has led to the introduction of the so-called radial distribution function, which is a special case of our second-order distribution functions. In the theory of the Brownian motion, another kind of multiform distribution function, representing the correlation of events at different times, has also been introduced quite recently (Ming Chen Wang & Uhlenbeck 1945).

In the present first part, we give only the definitions and fundamental equations of the theory, but we indicate how Boltzmann's gas theory follows as a special case, and how Kirkwood's derivation of the radial distribution function for 'rigid spherical molecules' (Kirkwood & Boggs 1942) can be generalized for an arbitrary potential between the molecules. A more detailed investigation of the integral equation obtained by this latter method will be left to Part II of this series. The equilibrium solution which results may then be made the foundation of a theory of viscosity, thermal conduction, self-diffusion, etc., for the liquid state.

2. NOTATION

We consider an assembly of N molecules confined within a volume V . We specify the position of the r th molecule by the vector $\mathbf{x}^{(r)}$, its velocity by $\boldsymbol{\xi}^{(r)}$, its acceleration by $\boldsymbol{\eta}^{(r)}$, its rate of acceleration by $\boldsymbol{\zeta}^{(r)}$, etc.

We denote by $n_h(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(h)}) \prod_{i=1}^h d\mathbf{x}^{(i)}$ the probability that h distinct molecules, which for convenience we number $1, 2, \dots, h$, occupy volume elements $d\mathbf{x}^{(1)}, d\mathbf{x}^{(2)}, \dots, d\mathbf{x}^{(h)}$, located at $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(h)}$, respectively at time t . We shall often abbreviate $n_h(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(h)})$ to $n_h(t, \mathbf{x})$, or even to n_h , the variables being sufficiently indicated by the suffix. In our notation, n_1 is simply the number density, and n_2/n_1^2 the radial distribution function of the molecular assembly in isotropic conditions.

Similarly, we denote by $f_h(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(h)}, \boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \dots, \boldsymbol{\xi}^{(h)}) \prod_{i=1}^h d\mathbf{x}^{(i)} d\boldsymbol{\xi}^{(i)}$ the probability that h distinct molecules, with their velocities in the ranges $\boldsymbol{\xi}^{(i)}, d\boldsymbol{\xi}^{(i)}$, $i = 1, 2, \dots, h$, occupy the volume elements $\mathbf{x}^{(i)}, d\mathbf{x}^{(i)}$, $i = 1, 2, \dots, h$, at time t . Then f_1 is the velocity distribution function commonly employed in the kinetic theory of gases. The f_h satisfy

$$\int \dots \int f_h(t, \mathbf{x}, \boldsymbol{\xi}) \prod_{i=1}^h d\boldsymbol{\xi}^{(i)} = n_h(t, \mathbf{x}). \quad (2.1)$$

In the same way an acceleration distribution function $g_h(t, \mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\eta})$ may be defined satisfying

$$\int \dots \int g_h(t, \mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\eta}) \prod_{i=1}^h d\boldsymbol{\eta}^{(i)} = f_h(t, \mathbf{x}, \boldsymbol{\xi}), \quad (2.2)$$

and, further, a rate-of-acceleration distribution function satisfying an analogous equation, etc.

There exist also formulae connecting distribution functions of the same kind, but with different values of h ; it is obvious, for example, from the definitions, that

$$\int n_{h+1}(t, \mathbf{x}) d\mathbf{x}^{(h+1)} = (N - h) n_h(t, \mathbf{x}), \quad (2.3)$$

$$\int \int f_{h+1}(t, \mathbf{x}, \boldsymbol{\xi}) d\mathbf{x}^{(h+1)} d\boldsymbol{\xi}^{(h+1)} = (N - h) f_h(t, \mathbf{x}, \boldsymbol{\xi}), \quad (2.4)$$

and similar relations must hold between the g_h 's and higher distribution functions.

3. THE EQUATIONS OF CONTINUITY

The equations of continuity are derived from the requirement that the probability of any molecule occupying the volume element $\mathbf{x}^{(i)}$, $d\mathbf{x}^{(i)}$ at time t must be the limit of the sum, over all values of $\xi^{(i)}$, of the probability of the same molecule having velocity $\xi^{(i)}$, and position $\mathbf{x}^{(i)} - \xi^{(i)}\delta t$, at time $t - \delta t$. Thus

$$\int \dots \int f_h(t - \delta t, \mathbf{x} - \xi \delta t, \xi) \prod_{i=1}^h d\xi^{(i)} = n_h(t, \mathbf{x}). \quad (3.1)$$

Subtracting from equation (2.1), we obtain

$$\int \dots \int \left(\frac{\partial f_h}{\partial t} + \sum_{i=1}^h \frac{\partial f_h}{\partial \mathbf{x}^{(i)}} \xi^{(i)} \right) \prod_{i=1}^h d\xi^{(i)} = 0, \quad (3.2)$$

or
$$\frac{\partial n_h}{\partial t} + \sum_{i=1}^h \frac{\partial}{\partial \mathbf{x}^{(i)}} (n_h \bar{\xi}^{(i)}) = 0, \quad (3.3)$$

where
$$\bar{\xi}^{(i)} = \frac{1}{n_h} \int \dots \int f_h \xi^{(i)} \prod_{i=1}^h d\xi^{(i)} \quad (3.4)$$

is the average velocity of the i th molecule in a group of h molecules whose positions are specified. For $h = 1$, (3.3) is the ordinary equation of continuity for a statistical assembly.

By precisely similar reasoning, we obtain for f_h ,

$$\frac{\partial f_h}{\partial t} + \sum_{i=1}^h \frac{\partial f_h}{\partial \mathbf{x}^{(i)}} \xi^{(i)} + \sum_{i=1}^h \frac{\partial}{\partial \xi^{(i)}} (f_h \bar{\eta}^{(i)}) = 0, \quad (3.5)$$

where
$$\bar{\eta}^{(i)} = \frac{1}{f_h} \int \dots \int g_h \eta^{(i)} \prod_{i=1}^h d\eta^{(i)} \quad (3.6)$$

is the mean acceleration of the i th molecule in a group of h molecules with given positions and velocities, the positions and velocities of the other molecules remaining unspecified.

Plainly g_h and the distribution functions for higher accelerations will all satisfy equations similar to (3.3) and (3.5).

4. THE EQUATIONS OF MOTION

Let Φ denote the potential energy of the entire assembly; we shall suppose, for convenience, that it can be expressed in the form

$$\Phi = \frac{1}{2} \sum_{r,s=1}^N \phi^{(rs)}, \quad (4.1)$$

where $\phi^{(rs)} = \phi(\mathbf{x}^{(r)}, \mathbf{x}^{(s)})$ is the mutual potential energy of two molecules at $\mathbf{x}^{(r)}$ and $\mathbf{x}^{(s)}$. When the intermolecular forces are central forces, $\phi^{(rs)}$ will depend only on

the distance $|\mathbf{x}^{(s)} - \mathbf{x}^{(r)}|$ between two molecules. In the sequel we assume this, and that the molecules are all alike; but there is always an easy generalization in case these conditions are not fulfilled.

The acceleration of a molecule at $\mathbf{x}^{(r)}$ is given by

$$\eta^{(r)} = -\frac{1}{m} \sum_{s=1}^N \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}} \quad (4.2)$$

in the absence of external forces, where m is the common mass of the molecules. Also

$$\zeta^{(r)} = -\frac{1}{m} \sum_{s,t=1}^N \left(\xi^{(t)} \frac{\partial}{\partial \mathbf{x}^{(t)}} \right) \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}}. \quad (4.3)$$

Hence we find the following expression for the mean acceleration $\overline{\eta^{(i)}}$, alternative to (3.6):

$$\overline{\eta^{(i)}} = -\frac{1}{m} \sum_{j=1}^h \frac{\partial \phi^{(ij)}}{\partial \mathbf{x}^{(i)}} - \frac{1}{m f_h} \iint \frac{\partial \phi^{(i,h+1)}}{\partial \mathbf{x}^{(i)}} f_{h+1} d\mathbf{x}^{(h+1)} d\xi^{(h+1)}. \quad (4.4)$$

When we substitute this expression in (3.5), we obtain

$$\frac{\partial f_h}{\partial t} + \sum_{i=1}^h \frac{\partial f_h}{\partial \mathbf{x}^{(i)}} \xi^{(i)} = \frac{1}{m} \sum_{i,j=1}^h \frac{\partial f_h}{\partial \xi^{(i)}} \frac{\partial \phi^{(ij)}}{\partial \mathbf{x}^{(i)}} + \frac{1}{m} \iint \sum_{i=1}^h \frac{\partial \phi^{(i,h+1)}}{\partial \mathbf{x}^{(i)}} \frac{\partial f_{h+1}}{\partial \xi^{(i)}} d\mathbf{x}^{(h+1)} d\xi^{(h+1)}. \quad (4.5)$$

By making use of (4.3) and similar formulae, equations similar to (4.5) are easily derived for g_h and the higher acceleration distribution functions.

The equations (4.5) for $h = 1, 2, \dots, N$ constitute a set of differential equations to determine the f_h , all except the last of which involve the solution of the next equation. The solution of these equations may therefore be effected either by commencing at the last equation and working backwards, or by using an approximate expression for f_{h+1} in terms of f_h . One such approximation results from the assumption that only binary encounters between the molecules need be considered, and the consequences of this assumption are examined in Appendix I, where it is shown that the equation (4.5) then reduces to the familiar Boltzmann equation of the kinetic theory of gases. Plainly a valid theory of the liquid state can never be founded on this particular basis, but there exists an alternative approximation which can be employed with far greater confidence. This is

$$f_3 = \frac{f_2^{(23)} f_2^{(31)} f_2^{(12)}}{f_1^{(1)} f_1^{(2)} f_1^{(3)}}, \quad (4.6)$$

expressing that the relative probability $\frac{f_3}{f_1^{(1)} f_1^{(2)} f_1^{(3)}}$ of the occurrence of a molecule $\mathbf{x}^{(3)}$, $\xi^{(3)}$ in conjunction with two others should be the product of the relative probabilities $\frac{f_2^{(31)}}{f_1^{(3)} f_1^{(1)}}$ and $\frac{f_2^{(32)}}{f_1^{(3)} f_1^{(2)}}$ of the occurrence of the same molecule in conjunction with each of the others separately.

5. THE EQUATIONS OF EQUILIBRIUM

From the fundamental equations (4.5) for the f_h , we now endeavour to find the equilibrium solution by substituting $\frac{\partial f_h}{\partial t} = 0$. It is known from statistical mechanics that the solution must be of the form

$$f_h = n_h \left(\frac{m}{2\pi kT} \right)^{\frac{3h}{2}} \exp \left\{ -\frac{m}{2kT} \sum_{i=1}^h \xi_i^2 \right\}. \quad (5.1)$$

By trial we find that (5.1) does in fact satisfy (4.5), provided

$$\frac{\partial n_h}{\partial \mathbf{x}^{(i)}} + \sum_{j=1}^h \frac{n_h}{kT} \frac{\partial \phi^{(ij)}}{\partial \mathbf{x}^{(i)}} + \int \frac{n_{h+1}}{kT} \frac{\partial \phi^{(i,h+1)}}{\partial \mathbf{x}^{(i)}} d\mathbf{x}^{(h+1)} = 0. \quad (5.2)$$

When $h = N$, the integrand of (5.2) vanishes, and the residual equation can be integrated, with the result

$$Q n_N = \exp(-\Phi/kT), \quad (5.3)$$

where Q is a constant of integration. This formula embodies a well-known result of statistical mechanics. By repeated use of (2.3) we find

$$Q = \frac{1}{N!} \int \dots \int e^{-\Phi/kT} \prod_{i=1}^N d\mathbf{x}^{(i)}, \quad (5.4)$$

so that Q is simply the phase integral which is the starting point of many investigations in statistical thermodynamics. From (5.3) we obtain also

$$Q n_h = \frac{1}{(N-h)!} \int \dots \int e^{-\Phi/kT} \prod_{i=h+1}^N d\mathbf{x}^{(i)}, \quad (5.5)$$

which completes the formal solution.

The evaluation of the integrals (5.4) and (5.5) is, however, so complicated for any but the simplest forms of potential that it seems preferable, for practical purposes, to make immediate use of (5.2) when h is small, in conjunction with some relation such as (4.6), which now reduces to

$$n_1^3 n_3 = n_2^{(12)} n_2^{(23)} n_2^{(31)}. \quad (5.6)$$

This is equivalent to the 'superposition approximation' which has been examined and used by Kirkwood & Boggs (1942). When (5.6) is substituted in equation (5.2), we find, after some analysis, which is given in Appendix II, that

$$\log \frac{n_2(r)}{n_1^2} + \frac{\phi(r)}{kT} = \frac{\pi}{n_1^3} \int_0^\infty \int_{-s}^s (s^2 - t^2) \frac{t+r}{r} \{n_2(t+r) - n_1^2\} dt n_2(s) \frac{\phi'(s)}{kT} ds, \quad (5.7)$$

where $n_2(r)$ and $\phi(r)$ depend only on the distance r between the two molecules. This is the generalization for an arbitrary potential of an equation obtained by Kirkwood & Boggs (1942), to which it reduces in the case of 'rigid spherical molecules'. It is capable of giving solutions which agree well with the experimental data obtained from the study of X-ray scattering by monatomic liquids.

Once $n_2(r)$ has been determined, all the equilibrium properties of the assembly follow, for, the internal energy E is given by*

$$E = \frac{1}{2} \left(\sum_{r,s=1}^N \overline{\phi^{(rs)}} + m \sum_{r=1}^N \overline{\xi^{(r)2}} \right) = \frac{1}{2} V \int_0^\infty n_2(r) \phi(r) 4\pi r^2 dr + \frac{3}{2} N k T, \quad (5.8)$$

and the free energy A by
$$\frac{\partial}{\partial T} \left(\frac{A}{T} \right) = - \frac{E}{T^2}, \quad (5.9)$$

together with a suitable limiting condition, such as that provided by the third law of thermodynamics. It is well known that all equilibrium properties can be derived from A , when expressed as a function of n_1 and the temperature T .

6. THE DYNAMICAL EQUATIONS

It has been seen how it is possible to obtain the equilibrium solution of the equations

$$\frac{\partial f_1^{(r)}}{\partial t} + \frac{\partial f_1^{(r)}}{\partial \mathbf{x}^{(r)}} \xi^{(r)} + \frac{\partial f_1^{(r)}}{\partial \xi^{(r)}} \mathbf{F}^{(r)} = \iint \frac{\partial f_2^{(rs)}}{\partial \xi^{(r)}} \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}} d\mathbf{x}^{(s)} d\xi^{(s)}, \quad (6.1)$$

$$\begin{aligned} \frac{\partial f_2^{(rs)}}{\partial t} + \frac{\partial f_2^{(rs)}}{\partial \mathbf{x}^{(r)}} \xi^{(r)} + \frac{\partial f_2^{(rs)}}{\partial \mathbf{x}^{(s)}} \xi^{(s)} + \frac{\partial f_2^{(rs)}}{\partial \xi^{(r)}} \left(\mathbf{F}^{(r)} - \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}} \right) + \frac{\partial f_2^{(rs)}}{\partial \xi^{(s)}} \left(\mathbf{F}^{(s)} - \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(s)}} \right) \\ = \iint \left(\frac{\partial f_3^{(rst)}}{\partial \xi^{(r)}} \frac{\partial \phi^{(rt)}}{\partial \mathbf{x}^{(r)}} + \frac{\partial f_3^{(rst)}}{\partial \xi^{(s)}} \frac{\partial \phi^{(st)}}{\partial \mathbf{x}^{(s)}} \right) d\mathbf{x}^{(t)} d\xi^{(t)}, \end{aligned} \quad (6.2)$$

with the aid of (4.6). It will be noticed that an additional term has been introduced into these equations to meet with the possibility of an external force $\mathbf{F}^{(r)}$ which is a function of the position $\mathbf{x}^{(r)}$.

To obtain the most general solution which does not differ greatly from the equilibrium solution, we replace

$$f_1^{(r)} \text{ by } f_1^{(r)}(1 + u_1^{(r)}), \quad f_2^{(rs)} \text{ by } f_2^{(rs)}(1 + u_2^{(rs)}), \quad (6.3)$$

so that $f_1^{(r)}$ and $f_2^{(rs)}$ now refer to the equilibrium solution, and neglect terms quadratic in the $u_1^{(r)}$ and $u_2^{(rs)}$. This procedure gives

$$\frac{\partial u_1^{(r)}}{\partial t} + \frac{\partial u_1^{(r)}}{\partial \mathbf{x}^{(r)}} \xi^{(r)} + \frac{\partial u_1^{(r)}}{\partial \xi^{(r)}} \mathbf{F}^{(r)} = \frac{1}{f_1^{(r)}} \frac{\partial}{\partial \xi^{(r)}} \iint f_2^{(rs)} u_2^{(rs)} \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}} d\mathbf{x}^{(s)} d\xi^{(s)}, \quad (6.4)$$

$$\begin{aligned} \frac{\partial u_2^{(rs)}}{\partial t} + \frac{\partial u_2^{(rs)}}{\partial \mathbf{x}^{(r)}} \xi^{(r)} + \frac{\partial u_2^{(rs)}}{\partial \mathbf{x}^{(s)}} \xi^{(s)} + \frac{\partial u_2^{(rs)}}{\partial \xi^{(r)}} \mathbf{D}^{(rs)} + \frac{\partial u_2^{(rs)}}{\partial \xi^{(s)}} \mathbf{D}^{(rs)} \\ = \frac{1}{f_2^{(rs)}} \frac{\partial}{\partial \xi^{(r)}} \iint f_3^{(rst)} (u_2^{(rt)} + u_2^{(st)} - u_1^{(r)} - u_1^{(s)} - u_1^{(t)}) \frac{\partial \phi}{\partial \mathbf{x}^{(r)}} d\mathbf{x}^{(t)} d\xi^{(t)} \\ + \frac{1}{f_2^{(rs)}} \frac{\partial}{\partial \xi^{(s)}} \iint f_3^{(rst)} (u_2^{(st)} + u_2^{(rt)} - u_1^{(s)} - u_1^{(r)} - u_1^{(t)}) \frac{\partial \phi}{\partial \mathbf{x}^{(s)}} d\mathbf{x}^{(t)} d\xi^{(t)}, \end{aligned} \quad (6.5)$$

where

$$-\mathbf{D}^{(rs)} = \frac{\partial \phi^{(rs)}}{\partial \mathbf{x}^{(r)}} + \iint f_3^{(rst)} \frac{\partial \phi^{(rt)}}{\partial \mathbf{x}^{(r)}} d\mathbf{x}^{(t)} d\xi^{(t)} - \mathbf{F}^{(r)}. \quad (6.6)$$

* Additional terms would, of course, be required if the molecules had internal degrees of freedom.

Equations (6.4) and (6.5) constitute a pair of simultaneous linear integro-differential equations for the determination of $u_1^{(r)}$ and $u_2^{(r)}$. Once solved with the appropriate boundary conditions, the dynamical properties of the assembly may be investigated in the usual way.

7. APPENDIX I

We shall prove here that when $h = 1$, our equation (4.5) reduces to Boltzmann's equation, in the case when only binary encounters need be considered. First it is necessary to obtain an expression for $f_2(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)})$.

Examine the orbits of two molecules which at time t have positions $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$, and velocities $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}$, and consider motion relative to the mass-centre of the two molecules, which must be in one plane. Let $\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}$ be their velocities before the encounter, and $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}$ their velocities after the encounter. Write $\mathbf{r} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)}$, $\boldsymbol{\rho} = \boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)}$, and denote by \mathbf{k} the vector joining the intersections of the initial and final asymptotes to the orbits of the two molecules.

Now, as the molecules move along their respective orbits, the configurational probability $f_2(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}) d\mathbf{x}^{(1)} d\mathbf{x}^{(2)} d\boldsymbol{\xi}^{(1)} d\boldsymbol{\xi}^{(2)}$ must remain unchanged; also the 'volume' in phase space, $d\mathbf{x}^{(1)} d\mathbf{x}^{(2)} d\boldsymbol{\xi}^{(1)} d\boldsymbol{\xi}^{(2)}$, is unaltered, by Liouville's theorem. Hence

$$f_2(t, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}) = f_1(\boldsymbol{\lambda}^{(1)}) f_1(\boldsymbol{\lambda}^{(2)}), \quad (7.1)$$

where $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(2)}$ are given in terms of $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \mathbf{r}$ and \mathbf{k} by the conservation equations

$$\left. \begin{aligned} \boldsymbol{\lambda}^{(1)} + \boldsymbol{\lambda}^{(2)} &= \boldsymbol{\xi}^{(1)} + \boldsymbol{\xi}^{(2)}, \\ \mathbf{k} \times (\boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)}) &= \mathbf{r} \times (\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)}), \\ \boldsymbol{\lambda}^{(1)2} + \boldsymbol{\lambda}^{(2)2} &= \boldsymbol{\xi}^{(1)2} + \boldsymbol{\xi}^{(2)2} + \frac{2\phi(r)}{m}. \end{aligned} \right\} \quad (7.2)$$

Hence, by some easy vector algebra,

$$\mathbf{k}^2(\boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)}) = \pm \left\{ \mathbf{k}^2 \boldsymbol{\rho}^2 + \frac{4\mathbf{k}^2 \phi(r)}{m} + (\mathbf{r} \cdot \boldsymbol{\rho})^2 - \mathbf{r}^2 \boldsymbol{\rho}^2 \right\}^{\frac{1}{2}} \mathbf{k} + (\mathbf{k} \cdot \mathbf{r}) \boldsymbol{\rho} - (\mathbf{k} \cdot \boldsymbol{\rho}) \mathbf{r}. \quad (7.3)$$

The ambiguity in sign is due to the fact that $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}$ also satisfy the equations (7.2). Hence from (7.3), when \mathbf{k} is fixed,

$$\left(\mathbf{r} \frac{\partial}{\partial \boldsymbol{\rho}} \right) (\boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)}) \frac{2\phi'(r)}{mr} = \left(\boldsymbol{\rho} \frac{\partial}{\partial \mathbf{r}} \right) (\boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)}). \quad (7.4)$$

It is now possible to evaluate the right-hand side of (4.5), which, with $h = 1$, becomes

$$\begin{aligned} \frac{1}{m} \iint \frac{\partial f_2}{\partial \boldsymbol{\xi}^{(1)}} \frac{\partial \phi}{\partial \mathbf{x}^{(1)}} d\mathbf{x}^{(2)} d\boldsymbol{\xi}^{(2)} &= \frac{1}{m} \iint \left(\frac{\partial f_2}{\partial \boldsymbol{\xi}^{(1)}} - \frac{\partial f_2}{\partial \boldsymbol{\xi}^{(2)}} \right) \frac{\partial \phi}{\partial \mathbf{x}^{(1)}} d\mathbf{x}^{(2)} d\boldsymbol{\xi}^{(2)} \\ &= \iint \frac{\phi'(r)}{mr} \left(\mathbf{r} \frac{\partial}{\partial \boldsymbol{\rho}} \right) (\boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)}) \left\{ f_1(\boldsymbol{\lambda}^{(1)}) \frac{\partial f_1(\boldsymbol{\lambda}^{(2)})}{\partial \boldsymbol{\lambda}^{(2)}} - f_1(\boldsymbol{\lambda}^{(2)}) \frac{\partial f_1(\boldsymbol{\lambda}^{(1)})}{\partial \boldsymbol{\lambda}^{(1)}} \right\} d\mathbf{x}^{(2)} d\boldsymbol{\xi}^{(2)}. \end{aligned} \quad (7.5)$$

Along a trajectory, where \mathbf{k} is constant, the integrand reduces to

$$\left(\rho \frac{\partial}{\partial \mathbf{r}}\right)\{f_1(\lambda^{(1)})f_1(\lambda^{(2)})\},$$

because of (7.4) and the relations $\frac{\partial}{\partial \mathbf{r}} \lambda^{(2)} = -\frac{\partial}{\partial \mathbf{r}} \lambda^{(1)} = \frac{1}{2} \frac{\partial}{\partial \mathbf{r}} (\lambda^{(2)} - \lambda^{(1)})$, which follow immediately from (7.2). The integration over $\mathbf{x}^{(2)}$ is now performed by integrating first along a trajectory, and then over all values of $\mathbf{b} = \mathbf{k} \times \rho / \rho$. It is easily seen that the entire available space is covered just once in this way. The result is

$$\iint \{f_1(\mu^{(1)})f_1(\mu^{(2)}) - f_1(\xi^{(1)})f_1(\xi^{(2)})\} \rho d\mathbf{b} d\xi^{(2)}, \quad (7.6)$$

in complete agreement with Boltzmann's well-known formula.

8. APPENDIX II

We give here the derivation of the integral equation (5.7) from equation (5.2) with $h = 2$, and (5.6) inserted, namely,

$$\frac{\partial n_2^{(12)}}{\partial \mathbf{x}^{(2)}} + \frac{n_2^{(12)}}{kT} \frac{\partial \phi^{(12)}}{\partial \mathbf{x}^{(2)}} = \frac{n_2^{(12)}}{n_1^2 kT} \int n_2^{(23)} n_2^{(31)} \frac{\partial \phi^{(23)}}{\partial \mathbf{x}^{(3)}} d\mathbf{x}^{(3)}. \quad (8.1)$$

Writing $\mathbf{r} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)}$, $\mathbf{s} = \mathbf{x}^{(3)} - \mathbf{x}^{(2)}$, $\mathbf{t} = \mathbf{x}^{(3)} - \mathbf{x}^{(1)}$, and $\mathbf{r} \cdot \mathbf{s} = rs \cos \theta$; then, after scalar multiplication by \mathbf{r} ,

$$n_2'(r) + n_2(r) \frac{\phi'(r)}{kT} = \frac{2\pi n_2(r)}{n_1^2} \int_0^\infty \int_0^\pi n_2(s) n_2(t) \frac{\phi'(s)}{kT} s^2 \sin \theta \cos \theta d\theta ds, \quad (8.2)$$

where $t = (r^2 + s^2 + 2rs \cos \theta)^{1/2}$, or

$$\frac{d}{dr} \left\{ \log n_2(r) + \frac{\phi(r)}{kT} \right\} = \frac{\pi}{n_1^2 r^2} \int_0^\infty \int_{|r-s|}^{|r+s|} n_2(t) t (t^2 - r^2 - s^2) dt n_2(s) \frac{\phi'(s)}{kT} ds. \quad (8.3)$$

We define $n_2(r)$ and $\phi(r)$ for negative r by the equations $n_2(r) = n_2(-r)$ and $\phi(r) = \phi(-r)$; it is then found that it is possible to integrate the right-hand side of (8.3) by parts, giving

$$\log n_2(r) + \frac{\phi(r)}{kT} = \frac{\pi}{n_1^2} \int_0^\infty \int_{-s}^s (s^2 - t^2) \frac{t+r}{r} n_2(t+r) dt n_2(s) \frac{\phi'(s)}{kT} ds + a, \quad (8.4)$$

where the constant of integration a is fixed by $n_2(r) \rightarrow n_1^2$ as $r \rightarrow \infty$. Thus

$$\log n_1^2 = \frac{\pi}{n_1^2} \int_0^\infty \int_{-s}^s (s^2 - t^2) \frac{t+r}{r} n_1^2 dt n_2(s) \frac{\phi'(s)}{kT} ds + a, \quad (8.5)$$

and, by subtraction from (8.4), (5.7) follows.

9. CONCLUSION

We have described a set of molecular distribution functions adequate for the formulation of the fundamental equations of a kinetic theory of liquids. From these equations we have shown that Boltzmann's theory of gases and some of the results of statistical mechanics can be obtained as special cases. We have indicated how our equations can be solved and applied to the solution of both static and dynamic problems concerning liquids.

REFERENCES

- Born, M. 1923 *Atomtheorie des festen Zustandes*. Leipzig: Teubner.
 Chapman, S. & Cowling, T. G. 1939 *The mathematical theory of non-uniform gases*. Camb. Univ. Press.
 Kirkwood, J. G. & Boggs, E. M. 1942 *J. Chem. Phys.* **10**, 394.
 Ming Chen Wang & Uhlenbeck, G. E. 1945 *Rev. Mod. Phys.* **17**, 323.
 Vlasov, A. 1945 *J. Phys.* **9**, 25, 130.

Some applications of the Riesz potential to the theory of the electromagnetic field and the meson field

By N. E. FREMBERG, *Institute of Mathematics, University of Lund*

(Communicated by P. A. M. Dirac, F.R.S.—Received 14 January 1946)

This paper contains some applications of the method of Marcel Riesz in the solution of normal hyperbolic differential equations, in particular the wave equation, where the known difficulties, due to the occurrence of divergent integrals, are avoided by a process of analytical continuation. In the theory of the electromagnetic field the method yields simple deductions of classical results, but also the results recently obtained by Dirac regarding the proper energy and proper momentum of an electron are obtained without any addition of new assumptions. The corresponding problem in Bhabha's analogous theory for the neutral meson field are also studied.

1. INTRODUCTION

By introducing a generalization of the Riemann-Liouville integral Professor M. Riesz (1936, 1937) has created a new method for treatment of normal hyperbolic differential equations, in particular the wave equation in an arbitrary number of dimensions. This generalization opens the widest possibilities for theoretical investigations as well as for practical adaptations. One advantage of the method is, for instance, that the well-known difficulties within the theory of hyperbolic differential equations which are due to the occurrence of divergent integrals and which were mastered by Hadamard (1932) by means of the important conception 'partie

finie' can be completely avoided. Another is that the equations in even and odd dimensions, differing in some respects ('Huygens's principle'), can be treated analogously, while Hadamard reduces one type on the other by a special 'méthode de descente'. Riesz gives the solution of the differential equation in the form of the analytical continuation of an expression which depends analytically on a parameter. Baker & Copson (1939) have applied Riesz's method to the cylindrical wave equation, and Copson (1943) has further published some applications of the method to other problems within mathematical physics. Recently, Gustafson (1945, 1946) has applied the method in quantum mechanics to the question of the divergencies in the theory of interaction of an electron with the electromagnetic field and of a nucleon with a meson field. I hope later to be able to return to a more extensive discussion of the method as applied to the general wave equation. In the present paper I only want to give some applications of it within the theory of the electromagnetic field and the meson field. § 2 gives a short survey of Riesz's method for the common wave equation and introduces the conception of Riesz potentials, § 3 describes its application to Maxwell's equations and deduces the classical retarded potential, § 4 studies the Riesz potential and the Riesz field strength in the field created by a single electron, where the classical results of Liénard and Wiechert are found to be verified. In § 5 the Lorentz-Dirac equation for the electron is deduced; among other things it is evident that a finite proper-energy and proper-momentum should be ascribed to the electron. The remaining sections, finally, treat the corresponding problems in Bhabha's (1939) theory of the neutral meson field.

2. RIESZ'S METHOD FOR THE WAVE EQUATION. RIESZ POTENTIALS

In the four-dimensional Lorentz space, where the velocity of light is $= 1$, V is a vector with the contravariant components V^μ ($\mu = 0, 1, 2, 3$) and length V , the square of which is measured by the Lorentz form

$$V^2 = V^0^2 - V^1^2 - V^2^2 - V^3^2 = g_{\mu\nu} V^\mu V^\nu = g^{\mu\nu} V_\mu V_\nu = V_\mu V^\mu \quad (\mu, \nu = 0, 1, 2, 3),$$

where $g_{\mu\nu}$ is the metrical fundamental tensor and the ordinary tensorial notation is used. The scalar product $V_\mu W^\mu$ of two arbitrary vectors V and W is in the following denoted by (V, W) , and the vector product with the contravariant components $V^\mu W^\nu - V^\nu W^\mu$ by $[V, W]$. A point P is determined by the position vector $x = (x^\mu)$. The vector $r_{QP} = x - z$ between two points $Q(z)$ and $P(x)$ has the length r_{QP} , hence

$$r_{QP}^2 = (x - z, x - z) = g_{\mu\nu} (x^\mu - z^\mu) (x^\nu - z^\nu).$$

If $r_{QP} = 0$ and P is fixed, Q lies on a cone, the so-called light cone, with its vertex in P . We are here only considering the retrograde cone, characterized by $x^0 - z^0 > 0$, and in the sequel denoted by D^P . We further assume that S is a space-like surface, i.e. a manifold of three dimensions with each tangent plane space-like in the meaning of the theory of relativity, with the equation $S(x^\mu) = 0$, where $S(x^\mu)$ is assumed to be continuous together with its partial derivatives of sufficiently high order; to

avoid non-essential complications they are not defined more closely. S borders together with D^P a domain, within which $r_{QP}^2 \geq 0$, and which we denote by D_S^P . The part of S which is cut out by D^P is denoted by S^P .

A function $f(x^\mu)$ may be defined in the whole Lorentz space. We have then to determine a function Φ , which will satisfy the wave equation

$$\square \Phi = g^{\mu\nu} \frac{\partial^2 \Phi}{\partial x^\mu \partial x^\nu} = f, \quad (1)$$

and which will, together with some derivative, say with respect to x^0 , assume the prescribed boundary values on the surface S . These values, as well as f , are to be sufficiently regular to enable us to perform the following operations.

We introduce with Riesz the four-dimensional integral of the Riemann-Liouville type

$$I^\alpha f(P) = \frac{1}{H(\alpha)} \int_{D_S^P} f(Q) r_{QP}^{\alpha-4} dQ, \quad (2)$$

where

$$H(\alpha) = 2^{\alpha-1} \pi \Gamma\left(\frac{\alpha}{2}\right) \Gamma\left(\frac{\alpha-2}{2}\right).$$

If f is continuous the integral converges for $\alpha > 2$ and is, by analytical continuation, defined for $\alpha \leq 2$. In particular the fundamental formula

$$I^0 f(P) = f(P)$$

is valid when f and S and their first derivatives are continuous. Further

$$\square I^{\alpha+2} = I^\alpha$$

is also valid.

We now introduce

$$I^{*\alpha} \overline{f, g, h}(P) = \frac{1}{H(\alpha)} \int_{D_S^P} f(Q) r_{QP}^{\alpha-4} dQ + \frac{1}{H(\alpha)} \int_{S^P} \left\{ g(Q) r_{QP}^{\alpha-4} - h(Q) \frac{\partial r_{QP}^{\alpha-4}}{\partial n} \right\} dS, \quad (3)$$

where n denotes the normal of S , directed inwards D_S^P , and dS the three-dimensional surface element of S , with all expressions determined and measured in the meaning of Lorentz. We then have

$$\square I^{*\alpha+2} = I^{*\alpha}$$

and

$$I^{*0} \overline{f, g, h}(P) = f(P)$$

in the sense of analytical continuation, and get the desired solution of (1) expressed as

$$\Phi(P) = I^{*2} \overline{f, \frac{\partial \Phi}{\partial n}, \Phi}(P) \quad (4)$$

in the meaning of the analytical continuation of the corresponding expression (3) to $\alpha = 2$. According to (3) and (4) Φ is expressed by three integrals, analogous to the potentials of mass distribution in the space, simple and double layer in the ordinary

potential theory. We call them in their general form (3) *Riesz potentials*, as well as the resulting expression $I^{\alpha} \overline{f, g, h}$.

The Riesz potential $\Phi^{(\alpha)}(P) = I^{\alpha} \overline{f, \frac{\partial \Phi}{\partial n}, \Phi}(P)$

associated to the present problem, is thus for $\alpha = 2$ the classical potential $\Phi^{(2)} = \Phi$. The analytical continuation down to $\alpha = 0$ has been performed earlier in an explicit way for the integral (2) (Fremberg 1945). The remaining parts are principally to be treated analogously and will be published in a later paper.

3. MAXWELL'S EQUATIONS FOR THE ELECTROMAGNETIC FIELD

The problem is considerably simplified, and we approach the subsequent applications if the boundary surface S is the infinitely remote plane $x^0 = -\infty$, i.e. if the integration domain is the whole space bounded by D^P , which we denote by $D_{-\infty}^P$. We presuppose that f has such properties that convergence difficulties do not arise by making the integration domain infinite.

We assume that $f = 4\pi \mathbf{s}$, where \mathbf{s} is an electric current density vector (4-vector), and obtain the Riesz vector-potential of the field

$$\Phi^{(\alpha)}(P) = \frac{4\pi}{H(\alpha)} \int_{D_{-\infty}^P} \mathbf{s}(Q) r_{QP}^{\alpha-4} dQ. \quad (5)$$

Then we have

$$\square \Phi^{(\alpha)}(P) = \Phi^{(\alpha-2)}.$$

If, further,

$$\operatorname{div} \mathbf{s} = \frac{\partial s^{\mu}}{\partial x^{\mu}} = 0$$

we get

$$\operatorname{div} \Phi^{(\alpha)} = 0,$$

i.e. the Riesz field strength $(\{F_{01}, F_{02}, F_{03}\} = \{E_x, E_y, E_z\}; \{F_{32}, F_{13}, F_{21}\} = \{H_x, H_y, H_z\})$

$$\mathbf{F}^{(\alpha)} = \operatorname{rot} \Phi^{(\alpha)} \left(F_{\mu\nu}^{(\alpha)} = \frac{\partial \Phi_{\nu}^{(\alpha)}}{\partial x^{\mu}} - \frac{\partial \Phi_{\mu}^{(\alpha)}}{\partial x^{\nu}} \right)$$

satisfies *Maxwell's equations*, if the corresponding Riesz current density is defined naturally as

$$\mathbf{s}^{(\alpha)} = \frac{1}{4\pi} \Phi^{(\alpha-2)}.$$

We then carry out the analytical continuation of (5) to $\alpha = 2$. We put

$$\left[\sum_{i=1}^3 (x^i - z^i)^2 \right]^{\frac{1}{2}} = \rho,$$

where, consequently, $0 \leq \rho \leq x^0 - z^0$, and introduce as integration variables $r = r_{QP}$, ξ^1 , ξ^2 , ξ^3 with

$$z^0 = x^0 - (r^2 + \rho^2)^{\frac{1}{2}}, \quad z^i = x^i + \xi^i \quad (i = 1, 2, 3),$$

which gives $dQ = dz^0 dz^1 dz^2 dz^3 = \frac{r}{\sqrt{(r^2 + \rho^2)}} dr d\xi^1 d\xi^2 d\xi^3$

and $\Phi^{(\alpha)}(\mathbf{x}) = \frac{\alpha-2}{2^{\alpha-2} \{\Gamma(\frac{1}{2}\alpha)\}^2} \int_0^\infty r^{\alpha-3} dr \iiint_V \frac{s(x^0 - \sqrt{(r^2 + \rho^2)}, x^i + \xi^i)}{\sqrt{(r^2 + \rho^2)}} d\xi^1 d\xi^2 d\xi^3,$

where the integration domain V is the whole three-dimensional space. Using the formula

$$\lim_{\beta \rightarrow +0} \beta \int_0^\infty f(x) x^{\beta-1} dx = f(0), \quad (6)$$

which is valid if the integral converges for some $\beta > 0$ and $f(x)$ is continuous at $x = 0$, we get

$$\Phi(\mathbf{x}) = \lim_{\alpha \rightarrow 2+0} \Phi^{(\alpha)}(\mathbf{x}) = \iiint_V \frac{s(x^0 - \rho, x^i + \xi^i)}{\rho} d\xi^1 d\xi^2 d\xi^3,$$

where $\rho = (\xi^{1^2} + \xi^{2^2} + \xi^{3^2})^{1/2}$, i.e. the *classical retarded potential*.*

4. THE ELECTROMAGNETIC FIELD AROUND ONE SINGLE ELECTRON

If the current generates from one single point-shaped particle, e.g. an electron with charge e , we obtain by transition to the limit or by direct attempt (the current is line-shaped, disappearing outside the (time-directed) world-line of the electron)

$$\Phi^{(\alpha)}(P) = \frac{4\pi e}{H(\alpha)} \int_{-\infty}^{\tau_0} \mathbf{u}(Q) r_{QP}^{\alpha-4} d\tau. \quad (7)$$

$P(\mathbf{x})$ is a point outside the world-line, $Q(\mathbf{z})$ a point on the world-line, and $\mathbf{z} = \mathbf{z}(\tau)$ the equation of the world-line. The parameter τ is the length of arc of the world-line (the *proper-time* of the electron), measured from some initial point and reckoned as positive in the direction of increasing z^0 . The upper limit of the integral corresponds to the point P_0 where the retrograde cone with its vertex in P intersects the world-line, the *retarded point* to P . Further $\mathbf{u} = d\mathbf{z}/d\tau$ is the 4-velocity of the electron ($\mathbf{u}^2 = 1$). Under the natural assumption that the mechanical velocity (3-velocity) always is less than a constant less than the velocity of light, the integral converges for $2 < \alpha < 3$. We seek its limit for $\alpha = 2 + 0$. As

$$\mathbf{u} = \frac{d\mathbf{z}}{d\tau} = -\frac{d\mathbf{r}}{d\tau} \quad (\mathbf{r} = \mathbf{r}_{QP})$$

and

$$\frac{d\mathbf{r}}{d\tau} = -\frac{(\mathbf{r}, \mathbf{u})}{r}, \quad (8)$$

we get

$$\Phi^{(\alpha)}(P) = \frac{e(\alpha-2)}{2^{\alpha-2} \{\Gamma(\frac{1}{2}\alpha)\}^2} \int_0^\infty \frac{\mathbf{u}}{(\mathbf{r}, \mathbf{u})} r^{\alpha-3} dr,$$

and, according to (6) for $\alpha \rightarrow 2 + 0$,

$$\Phi(P) = \left\{ \frac{e\mathbf{u}}{(\mathbf{r}, \mathbf{u})} \right\}_0,$$

* Cf. Copson, p. 263.

where index 0 implies that the value of the function should be taken in the point retarded to P . We find here the *classical Liénard-Wiechert potential*.

In the sequel we assume that the world-line has such regularity properties that it is possible to perform the following operations, viz. existence and continuity of a sufficient number of derivatives of $\mathbf{z}(\tau)$ and that the world-line fulfils some natural conditions for $\tau \rightarrow -\infty$. We assume for example that the world-line for $\tau \rightarrow -\infty$ has a time-directed asymptote and that $\mathbf{z}(\tau)$ for $\tau < -T$ can be written

$$\mathbf{z}(\tau) = \mathbf{u}_{-\infty}\tau + \sum_{k=0}^{\infty} \lambda_k \tau^{-k} \quad (\mathbf{u}_{-\infty} = \lim_{\tau \rightarrow -\infty} \mathbf{u}).$$

To get the Riesz field strength we introduce the gradient operator ∇ with the covariant components $\nabla_\mu = \partial/\partial x^\mu$ and have

$$\mathbf{F}^{(\alpha)} = \text{rot } \Phi^{(\alpha)} = [\nabla, \Phi^{(\alpha)}].$$

As the integrand in (7) is singular in τ_0 , the expression (7) for $\Phi^{(\alpha)}$ is unsuitable for direct derivation. We use the relation $\mathbf{u} d\tau = d\mathbf{z}$ and write

$$\Phi^{(\alpha)} = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} d\mathbf{z} = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} \frac{\partial \mathbf{z}}{\partial r} dr,$$

where \mathbf{z} may be regarded as function of \mathbf{x} and r , chosen as independent variables. Thus we have

$$\begin{aligned} \nabla_\mu \Phi^{(\alpha)} &= \frac{\partial \Phi^{(\alpha)}}{\partial x^\mu} = \frac{4\pi e}{H(\alpha)} \int_{-\infty}^{\tau_0} r^{\alpha-4} \frac{\partial}{\partial x^\mu} \left(\frac{\partial \mathbf{z}}{\partial r} \right) dr \\ &= \frac{4\pi e}{H(\alpha)} \int_{-\infty}^{\tau_0} r^{\alpha-4} \frac{\partial}{\partial r} \left(\frac{\partial \mathbf{z}}{\partial x^\mu} \right) dr = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} d \left(\frac{\partial \mathbf{z}}{\partial x^\mu} \right). \end{aligned}$$

In order to get $\partial \mathbf{z} / \partial x^\mu$ we differentiate $r^2 = (\mathbf{x} - \mathbf{z}, \mathbf{x} - \mathbf{z})$ with respect to x^μ , which gives

$$0 = \frac{1}{2} \frac{\partial r^2}{\partial x^\mu} = r_\mu - \left(\mathbf{r}, \frac{\partial \mathbf{z}}{\partial x^\mu} \right) = r_\mu - (\mathbf{r}, \mathbf{u}) \frac{\partial \tau}{\partial x^\mu},$$

where r_μ is a covariant component of \mathbf{r} , $r_\mu = g_{\mu\nu} r^\nu = g_{\mu\nu} (x^\nu - z^\nu)$. From this we have

$$\frac{\partial \tau}{\partial x^\mu} = \frac{r_\mu}{(\mathbf{r}, \mathbf{u})},$$

and thus

$$\frac{\partial \mathbf{z}}{\partial x^\mu} = \frac{r_\mu \mathbf{u}}{(\mathbf{r}, \mathbf{u})}.$$

We then get, finally,

$$\nabla_\mu \Phi^{(\alpha)} = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} d \left(\frac{r_\mu \mathbf{u}}{(\mathbf{r}, \mathbf{u})} \right). \quad (9)$$

This gives

$$\mathbf{F}^{(\alpha)} = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} d \left(\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right), \quad (10)$$

which converges for $2 < \alpha < 5$. After partial integration $\mathbf{F}^{(\alpha)}$ takes the simple form

$$\mathbf{F}^{(\alpha)} = \frac{4\pi e(\alpha-4)}{H(\alpha)} \int_{-\infty}^{\tau_0} [\mathbf{r}, \mathbf{u}] r^{\alpha-6} d\tau, \quad (11)$$

which converges for $4 < \alpha < 5$. This is the same result which is obtained if $\Phi^{(\alpha)}$ in the form (7) is differentiated formally under the sign of integration.

Taking the limit of (10) for $\alpha = 2 + 0$ we get the classical Liénard-Wiechert field strength

$$\begin{aligned} \mathbf{F} = \mathbf{F}^{(2)} &= e \left\{ -\frac{d}{r dr} \left(\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right) \right\}_0 = e \left\{ \frac{1}{(\mathbf{r}, \mathbf{u})} \frac{d}{d\tau} \left(\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right) \right\}_0 \\ &= e \left\{ \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})^3} (1 - (\mathbf{r}, \dot{\mathbf{u}})) + \frac{[\mathbf{r}, \dot{\mathbf{u}}]}{(\mathbf{r}, \mathbf{u})^2} \right\}_0, \end{aligned}$$

where $\dot{\mathbf{u}} = d\mathbf{u}/d\tau$ = the 4-acceleration-vector.

$$\text{Outside the world-line} \quad \text{div } \Phi^{(\alpha)} = 0 \quad (12)$$

$$\text{and} \quad \square \Phi^{(\alpha)} = \Phi^{(\alpha-2)} \quad (13)$$

are valid, i.e. the quantities introduced satisfy Maxwell's equations, if the corresponding Riesz current density is defined as

$$\mathbf{s}^{(\alpha)} = \frac{1}{4\pi} \Phi^{(\alpha-2)}.$$

(12) follows immediately from (9), as

$$\text{div } \Phi^{(\alpha)} = (\nabla, \Phi^{(\alpha)}) = \frac{4\pi e}{H(\alpha)} \int_{\tau=-\infty}^{\tau_0} r^{\alpha-4} d \left(\frac{(\mathbf{r}, \mathbf{u})}{(\mathbf{r}, \mathbf{u})} \right) = 0,$$

and (13) follows from (7) by directly performing the operation \square under the sign of integration and considering that

$$\square \frac{r^{\alpha-4}}{H(\alpha)} = \frac{r^{\alpha-6}}{H(\alpha-2)}.$$

That this formal derivation under the sign of integration really yields a correct result is proved by partial integrations completely analogous to the computation of $\mathbf{F}^{(\alpha)}$ given above.*

5. THE INFLUENCE OF THE ELECTROMAGNETIC FIELD ON THE GENERATING ELECTRON ITSELF. THE PROPER LORENTZ FORCE OF THE ELECTRON. LORENTZ-DIRAC'S EQUATIONS OF MOTION

The applications of Riesz's method given above have only been very simplified deductions of classical results. It is, however, to be expected that this method may, by allowing us to avoid divergence and singularity difficulties, prove valuable for

* Note added May 1946. The above analysis becomes clearer if the differentiations with respect to x^μ are expressed by means of differentiations with respect to r or rather r^4 . This fact has kindly been pointed out to me by Professor Riesz, and I make extensive use of this device in my thesis which is to appear soon.

the solution of such problems where this kind of difficulties has as yet not been mastered, or at least only by means of artifices. One of these problems concerns the question of the proper-energy and proper-momentum of the electron.

As the Riesz quantities satisfy Maxwell's equations we may form the corresponding stress tensor $T^{(\alpha)}$ according to the usual scheme with the components

$$T^{(\alpha)\mu\rho} = \frac{1}{4\pi} \{ F^{(\alpha)\mu\nu} F^{(\alpha)\rho}_{\nu} + \frac{1}{2} g^{\mu\rho} F^{(\alpha)\kappa\lambda} F^{(\alpha)}_{\kappa\lambda} \}.$$

If $f^{(\alpha)}$ denotes the Riesz Lorentz force density we have

$$f^{(\alpha)\mu} = F^{(\alpha)\mu\nu} g^{(\alpha)}_{\nu}$$

and
$$f^{(\alpha)} = -\operatorname{div} T^{(\alpha)} \quad \text{or} \quad f^{(\alpha)\mu} = -\frac{\partial T^{(\alpha)\mu\rho}}{\partial x^{\rho}}.$$

According to Gauss's theorem the integral of $f^{(\alpha)}$ over a four-dimensional space is equal to the flow of energy-momentum through the (three-dimensional) boundary surface of the domain. This relation remains in the form of integrals convergent for $\alpha > 3$ and then, by analytical continuation, for $\alpha = 2$. The analytical continuation of the Lorentz force density integral over a region of four-dimensional space which surrounds a part of the world-line is the finite value*

$$\int_{\tau_1}^{\tau_2} e \bar{F}^{\mu\sigma} u_{\sigma} d\tau \quad \text{with} \quad \bar{F} = \frac{2e}{3} [\ddot{u}, u], \quad (14)$$

where the integral is to be taken along that part of the world-line which falls inside the domain in question. *Consequently, the analytical continuation of the energy-momentum flow is finite, too.*

(14) can now be interpreted in an interesting way. The current density vector $s^{(\alpha)}$ vanishes uniformly in every closed region outside the world-line for $\alpha = 2$. The integral of $s^{(\alpha)}$ over a four-dimensional region has the analytical continuation

$$\int_{\tau_1}^{\tau_2} e u d\tau$$

extended over the enclosed part of the world-line. The analytical continuation of the integral of $s^{(\alpha)}$ over a three-dimensional domain, lying in the plane orthogonal to the world-line in a point Q and having Q as an inner point, equals $e u$, i.e. the *convective current* generated by the particle, as is indeed very natural. Thus we are led to regard the integrand in (14) as the Lorentz force on the world-line, i.e. as the *proper Lorentz force* of the electron.

This interpretation is further corroborated by computing the analytical continuation of the field strength $F^{(\alpha)}$ in a point P on the world-line. We consider $F^{(\alpha)}$

* This has been proved although not published by M. Riesz. A simplified demonstration will be published later by the author of the present paper.

in the form (11). The integrand is divided into two parts, one corresponding to the interval $-\infty$ to $\tau_0 - \epsilon$, the other to $\tau_0 - \epsilon$ to τ_0 . The first integral converges for $\alpha < 5$, thus especially for $\alpha = 2$, and on account of the Γ -factor its contribution disappears for $\alpha = 2$. In the second integral the integrand develops in series. We denote for simplicity $\tau_0 = 0$ and get

$$\mathbf{r} = -\mathbf{u}_0 \tau - \dot{\mathbf{u}}_0 \frac{\tau^2}{2} - \ddot{\mathbf{u}}_0 \frac{\tau^3}{6} + O(\tau^4),$$

$$\mathbf{u} = \mathbf{u}_0 + \dot{\mathbf{u}}_0 \tau + \ddot{\mathbf{u}}_0 \frac{\tau^2}{2} + O(\tau^3),$$

$$r = -\tau + O(\tau^3),$$

$$[\mathbf{r}, \mathbf{u}] = -\frac{\tau^3}{2} \{[\mathbf{u}, \dot{\mathbf{u}}]_0 - \frac{2}{3}[\ddot{\mathbf{u}}, \mathbf{u}]_0 \tau\} + O(\tau^4).$$

Hence

$$\begin{aligned} & \frac{4\pi e(\alpha - 4)}{H(\alpha)} \int_{-\epsilon}^0 [\mathbf{r}, \mathbf{u}] r^{\alpha-6} d\tau \\ &= \frac{(4-\alpha)e}{2^{\alpha-1} \{\Gamma(\frac{1}{2}\alpha)\}^2} (\alpha-2) \int_{-\epsilon}^0 |\tau|^{\alpha-4} \{[\mathbf{u}, \dot{\mathbf{u}}]_0 - \frac{2}{3}[\ddot{\mathbf{u}}, \mathbf{u}]_0 \tau + O(\tau^2)\} d\tau, \end{aligned}$$

and the analytical continuation of $\mathbf{F}^{(\alpha)}$ thus becomes

$$\mathbf{F}^{(2)} = \mathbf{F} = \frac{2e}{3} [\ddot{\mathbf{u}}, \mathbf{u}]_0 (= \mathbf{F}). \quad (15)$$

From this we obtain the Lorentz force of the electron

$$f^\mu = F^{\mu\sigma} e u_\sigma = \frac{2e^2}{3} (\ddot{u}^\mu + \dot{\mathbf{u}}^2 u^\mu),$$

and, finally, the equations of motion for the electron

$$m\dot{\mathbf{u}} = \mathbf{f} \quad \text{or} \quad m\dot{u}^\mu = \frac{2e^2}{3} (\ddot{u}^\mu + \dot{\mathbf{u}}^2 u^\mu).$$

These modified Lorentz equations* were deduced by Dirac (1938) by an intuitive method, based upon sound physical considerations. Previously Lorentz got them as approximative equations.

We observe that the analytical continuation of the Riesz field strength outside the world-line gives the classical value but that it also gives a finite and very natural value on the world-line where the classical field strength has no meaning. The expression (15) is identical with the limit of $\frac{1}{2}(\mathbf{F}_{\text{ret.}} - \mathbf{F}_{\text{adv.}})$ in Dirac's work, if the point moves orthogonally to the world-line.

* The terms arising from the incident field do not cause any difficulties and are omitted in the developments above.

6. THE MESON FIELD

We consider Yukawa's equation of the meson field

$$(\square + \chi^2) u = f. \quad (16)$$

To the operator $\square + \chi^2$ we form a function $W^\alpha(Q, P)$ which plays the same part as the function $V^\alpha(Q, P) = \frac{1}{H(\alpha)} r_{QP}^{\alpha-4}$ in relation to the operator \square . We get

$$W^\alpha(Q, P) = \frac{1}{2^{k(\alpha+2)} \pi \Gamma(\frac{1}{2}\alpha)} \left(\frac{r_{QP}}{\chi} \right)^{k(\alpha-4)} J_{k(\alpha-4)}(\chi r_{QP}),$$

where J is a Bessel function.* Then we have

$$(\square + \chi^2) W^{\alpha+2} = W^\alpha,$$

and in particular

$$(\square + \chi^2) W^2 = 0,$$

where W^2 is the analytical continuation of W^α for $\alpha = 2$.

If we form the integral

$$I_M^\alpha f(P) = \int_{D_s^r} f(Q) W^\alpha(Q, P) dQ$$

analogous to the earlier integral $I^\alpha f(P)$, we can show that I_M^α fulfils the same fundamental relations as I^α . In particular

$$I_M^0 f(P) = f(P)$$

is valid.

If the function f is defined in the whole space the corresponding integral becomes

$$I_M^\alpha f(P) = \int_{D_\infty^r} f(Q) W^\alpha(Q, P) dQ.$$

By analytical continuation to $\alpha = 2$ we see that

$$u(P) = I_M^2 f(P)$$

is a solution of the equation

$$(\square + \chi^2) u = f,$$

such that $u(P)$ and $\partial u(P)/\partial x^0$ vanish when $x^0 \rightarrow -\infty$.

We consider with Bhabha (1939) the case where the field is created by one single neutron with the charge g , which describes its world-line with the velocity \mathbf{u} . In the sequel the same assumptions about the regularity of the world-line as those made earlier regarding the electron are supposed to be satisfied.

The Riesz potential for this field is ($r = r_{QP}$)

$$\Phi_M^{(g)}(P) = \frac{\chi^{k(4-\alpha)} g}{2^{k(\alpha-2)} \Gamma(\frac{1}{2}\alpha)} \int_{-\infty}^{\tau_0} \mathbf{u} r^{k(\alpha-4)} J_{k(\alpha-4)}(\chi r) d\tau \quad (17)$$

* Cf. Riesz (1937, p. 169).

which converges for $2 < \alpha < 3$. According to (8) we can write

$$\Phi_M^{(\alpha)}(P) = \frac{\chi^{k(4-\alpha)}g}{2^{k(\alpha-2)}\Gamma(\frac{1}{2}\alpha)} \int_0^\infty \frac{\mathbf{u}}{(\mathbf{r}, \mathbf{u})} r^{k(\alpha-2)} J_{k(\alpha-2)}(\chi r) dr.$$

Using the formula $J_{n-1}(x) = \frac{2n}{x} J_n(x) - J_{n+1}(x)$, we get

$$\begin{aligned} \Phi_M^{(\alpha)} = \frac{\chi^{k(4-\alpha)}g}{2^{k(\alpha-2)}\Gamma(\frac{1}{2}\alpha)} (\alpha-2) \int_0^\infty \frac{\mathbf{u}}{(\mathbf{r}, \mathbf{u})} r^{k(\alpha-4)} J_{k(\alpha-2)}(\chi r) dr \\ - \frac{\chi^{k(4-\alpha)}}{2^{k(\alpha-2)}\Gamma(\frac{1}{2}\alpha)} \int_0^\infty \frac{\mathbf{u}}{(\mathbf{r}, \mathbf{u})} r^{k(\alpha-2)} J_{k(\alpha)}(\chi r) dr. \end{aligned} \quad (18)$$

To get the limit of this expression for $\alpha = 2$ we apply to the first term a slight modification of (6), which is still true if $f(x)$ depends on β , $f(x) = g(x, \beta)$, and $g(x, \beta)$ is continuous in $x = \beta = 0$, and $f(0)$ is replaced by $g(0, 0)$. Apart from a factor the first term of (18) can be written

$$(\alpha-2) \int_0^\infty r^{\alpha-3} \frac{\mathbf{u}}{(\mathbf{r}, \mathbf{u})} r^{k(2-\alpha)} J_{k(\alpha-2)}(\chi r) dr$$

where the function $r^{k(2-\alpha)} J_{k(\alpha-2)}(\chi r)$ is continuous in $r = 0$, $\alpha = 2$. We thus get the potential of the meson field*

$$\Phi_M^{(2)} = \Phi_M = \left\{ \frac{g\mathbf{u}}{(\mathbf{r}, \mathbf{u})} \right\}_0 - \chi g \int_{-\infty}^{\tau_0} \frac{\mathbf{u}}{r} J_1(\chi r) dr,$$

where the first term of the right-hand member is the usual retarded potential, to which Φ_M reduces when $\chi \rightarrow 0$.

We try to find the potential of the meson field when the motion of the neutron is uniform. Then \mathbf{u} is constant and we get

$$\Phi_M^{(\alpha)} = \frac{\chi^{k(4-\alpha)}g\mathbf{u}}{2^{k(\alpha-2)}\Gamma(\frac{1}{2}\alpha)} \int_0^\infty \frac{r^{k(\alpha-2)}}{(\mathbf{r}, \mathbf{u})} J_{k(\alpha-2)}(\chi r) dr.$$

The integral, denoted by A , is easily calculated by transferring the world-line by means of a Lorentz transformation to the x^0 -axis. At this r and (\mathbf{r}, \mathbf{u}) remain invariant. Further we have $(\mathbf{r}, \mathbf{u}) = (r^2 + \rho^2)^{\frac{1}{2}}$, where $\rho = (\mathbf{r}, \mathbf{u})_0$. Hence

$$A = \int_0^\infty \frac{r^{k(\alpha-2)}}{\sqrt{(r^2 + \rho^2)}} J_{k(\alpha-2)}(\chi r) dr$$

and, by virtue of the formula (see e.g. Watson 1922),

$$\int_0^\infty \frac{x^{\nu+1}}{(x^2 + k^2)^{\mu+1}} J_\nu(ax) dx = \frac{a^\mu k^{\nu-\mu}}{2^\mu \Gamma(\mu+1)} K_{\nu-\mu}(ak),$$

we get

$$A = \sqrt{\frac{2}{\chi\pi}} \rho^{k(\alpha-2)} K_{k(\alpha-2)}(\chi\rho),$$

* Cf. Bhabha (1939, p. 388, formula (9)).

so that, finally, $\Phi_M^{(\alpha)} = \frac{\chi^{4(3-\alpha)} g \mathbf{u}}{2^{4(\alpha-3)} \pi^4 \Gamma(\frac{1}{2}\alpha)} \rho^{4(\alpha-3)} K_{4(\alpha-3)}(\chi \rho).$

As $K_{-\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^{-x}$ we get for $\alpha = 2$

$$\Phi_M^{(2)} = \Phi_M = \frac{g \mathbf{u}}{\rho} e^{-\chi \rho}$$

or

$$\Phi_M = \left\{ \frac{g \mathbf{u}}{(\mathbf{r}, \mathbf{u})} e^{-\chi(\mathbf{r}, \mathbf{u})} \right\}_0.$$

Further, we form the Riesz field strength $\mathbf{F}_M^{(\alpha)}$. Using the formula

$$\frac{d}{dx} (x^n J_n(x)) = x^n J_{n-1}(x)$$

we obtain by formal differentiation of (17) under the sign of integration (that this is permissible can be verified analogously to the corresponding calculation for the electromagnetic field of the electron)

$$\mathbf{F}_M^{(\alpha)} = \text{rot } \Phi_M^{(\alpha)} = [\nabla, \Phi_M^{(\alpha)}] = \frac{\chi^{4-\alpha} g}{2^{4(\alpha-2)} \Gamma(\frac{1}{2}\alpha)} \int_{-\infty}^{\tau_0} [\mathbf{r}, \mathbf{u}] (\chi r)^{4(\alpha-6)} J_{4(\alpha-6)}(\chi r) d\tau, \quad (19)$$

which converges for $4 < \alpha < 5$. Now

$$\begin{aligned} \int_{-\infty}^{\tau_0} [\mathbf{r}, \mathbf{u}] (\chi r)^{4(\alpha-6)} J_{4(\alpha-6)}(\chi r) d\tau &= \frac{1}{\chi} \int_0^{\infty} \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} (\chi r)^{4(\alpha-4)} J_{4(\alpha-6)}(\chi r) dr \\ &= \frac{1}{\chi^2} \int_{\tau=-\infty}^{\tau_0} \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} d\{(\chi r)^{4(\alpha-4)} J_{4(\alpha-4)}(\chi r)\} \\ &= -\frac{1}{\chi^2} \left[\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} (\chi r)^{4(\alpha-4)} J_{4(\alpha-4)}(\chi r) \right]_{\tau=-\infty}^{\tau_0} + \frac{1}{\chi^2} \int_{-\infty}^{\tau_0} \frac{d}{d\tau} \left\{ \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right\} (\chi r)^{4(\alpha-4)} J_{4(\alpha-4)}(\chi r) d\tau. \end{aligned}$$

The first term of the last member disappears for $4 < \alpha < 5$, and we get

$$\mathbf{F}_M^{(\alpha)} = \frac{\chi^{4-\alpha} g}{2^{4(\alpha-2)} \Gamma(\frac{1}{2}\alpha)} \int_{-\infty}^{\tau_0} \frac{d}{d\tau} \left\{ \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right\} (\chi r)^{4(\alpha-4)} J_{4(\alpha-4)}(\chi r) d\tau.$$

This integral is of the same type as (17). The analytical continuation to $\alpha = 2$ becomes

$$\mathbf{F}_M^{(2)} = \mathbf{F} = \left\{ \frac{g}{(\mathbf{r}, \mathbf{u})} \frac{d}{d\tau} \left(\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right) \right\}_0 - g \chi \int_{-\infty}^{\tau_0} \frac{d}{d\tau} \left\{ \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right\} \frac{J_1(\chi r)}{r} d\tau,$$

or, after partial integration of the second term of the right-hand member and using the formulae

$$\lim_{x \rightarrow 0} \frac{1}{x} J_1(x) = \frac{1}{2} \quad \text{and} \quad \frac{d}{dx} (x^n J_n(x)) = -x^n J_{n+1}(x),$$

$$\mathbf{F}_M = \left\{ \frac{g}{(\mathbf{r}, \mathbf{u})} \frac{d}{d\tau} \left(\frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right) \right\}_0 - \frac{g \chi^2}{2} \left\{ \frac{[\mathbf{r}, \mathbf{u}]}{(\mathbf{r}, \mathbf{u})} \right\}_0 + g \chi^2 \int_{-\infty}^{\tau_0} [\mathbf{r}, \mathbf{u}] \frac{J_2(\chi r)}{r^2} d\tau.$$

Here the first term is the retarded field strength to which the meson field degenerates when $\chi \rightarrow 0$.*

* Cf. Bhabha (1939, p. 388, formulae (11), (12), (13)). Our \mathbf{F}_M is $-G$ in Bhabha.

7. THE REACTION OF THE MESON FIELD ON THE HEAVY PARTICLE ITSELF. THE MESON FIELD ENERGY AND MOMENTUM FOR A PARTICLE IN UNIFORM MOTION

The same reasoning as was applied to the electromagnetic field in § 5 can here be applied to the meson field. For the moment we confine ourselves to calculating the proper field strength of the neutron, i.e. the analytical continuation of (19) for $\alpha = 2$ when the point P is in the world-line. After computations analogous to those performed above we obtain the *proper field strength*

$$\mathbf{F}_M = \frac{2g}{3} [\ddot{\mathbf{u}}, \mathbf{u}]_0 + g\chi^2 \int_{-\infty}^{r_0} [\mathbf{r}, \mathbf{u}] \frac{J_2(\chi r)}{r^2} d\tau$$

from which follow Bhabha's equations of motion (in the absence of external field)

$$M\dot{u}^\mu = gF_M^{\mu\nu}u_\nu,$$

where M is the mass of the neutron.

As the Riesz quantities are easily seen to satisfy the equations of the meson field*

$$\frac{\partial F_M^{(\alpha)\mu\nu}}{\partial x^\nu} - \chi^2 \Phi_M^{(\alpha)\mu} = -4\pi s_M^{(\alpha)\mu}$$

if the corresponding charge current density is defined as $s_M^{(\alpha)} = \frac{1}{4\pi} \Phi_M^{(\alpha-2)}$, the usual formalism of forming other field quantities can be applied, in particular to form the stress tensor $T_M^{(\alpha)}$ with the components†

$$T_M^{(\alpha)\mu\nu} = \frac{1}{4\pi} \left\{ F_M^{(\alpha)\mu\sigma} F_M^{(\alpha)\nu} + \frac{1}{2} g^{\mu\nu} F_M^{(\alpha)\sigma\rho} F_M^{(\alpha)\sigma\rho} + \chi^2 \Phi_M^{(\alpha)\mu} \Phi_M^{(\alpha)\nu} - \frac{\chi^2}{2} g^{\mu\nu} \Phi_M^{(\alpha)} \Phi_M^{(\alpha)} \right\}.$$

We calculate the analytical continuation of $T_M^{(\alpha)\mu\nu}$ over the total three-dimensional space $x^0 = \text{constant}$ (i.e. the flow of energy and momentum through the plane $x^0 = \text{const.}$) in case of a neutron in uniform motion. After some computations we get as value of the total energy and momentum

$$\text{anal. cont.}_{(\alpha=2)} \int_{(x^0=\text{const.})} T_M^{(\alpha)\mu\nu} dV = -\frac{1}{2} g^2 \chi u^\mu.$$

In the special case of a neutron *at rest* we have here the expression $-\frac{1}{2} g^2 \chi$ for the meson field energy which Bhabha obtains after subtracting a singular term.

In terminating this paper I want to express to Professor Marcel Riesz my sincerest gratitude for his direction of my attention to these problems, for valuable discussions and kind encouragement, and for critical examination of the manuscript. I am very grateful to Dr K.-G. Hagstroem and Dr Lars Gårding for careful examination of the manuscript.

* Cf. Bhabha (1939, p. 386). We have supposed $S^{12} = 0$.

† Cf. Bhabha (1939, p. 389, formula (14)).

REFERENCES

- Baker, B. B. & Copson, E. T. 1939 *The mathematical theory of Huygens' principle*, pp. 54-67. Oxford Univ. Press.
- Bhabha, H. F. 1939 *Proc. Roy. Soc. A*, **172**, 384-409.
- Copson, E. T. 1943 *Proc. Roy. Soc. Edinb. A*, **61**, 260-272.
- Dirac, P. A. M. 1938 *Proc. Roy. Soc. A*, **167**, 148-169.
- Fremberg, N. E. 1945 *K. fysiogr. Sällsk. Lund Förh.* **15**, no. 27, pp. 1-12.
- Gustafson, T. 1945 *K. fysiogr. Sällsk. Lund Förh.* **15**, no. 28, pp. 1-12.
- Gustafson, T. 1946 *K. fysiogr. Sällsk. Lund Förh.* **16**, no. 2, pp. 1-9.
- Hadamard, F. 1932 *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*. Paris.
- Riesz, M. 1936 *C.R. Int. Congr. Math. Oslo*, ii, pp. 44-45.
- Riesz, M. 1937 *Société mathématique de France, Conférences de la Réunion internationale des mathématiciens tenue à Paris en juillet 1937*, pp. 153-170 (Paris, 1939).
- Watson 1922 *Theory of Bessel function*, p. 434. Cambridge Univ. Press.

The crystal structure of some molecular complexes of 4:4'-dinitrodiphenyl

I. The complex with 4-hydroxydiphenyl

BY D. H. SAUNDER, M.Sc.

University of Cape Town (Beit Railway Trust Rhodesian Fellow)

(Communicated by Sir Laurence Bragg, F.R.S.—Received 22 January 1946)

The method of double Fourier series has been applied to determine the crystal structure of the complex between 4:4'-dinitrodiphenyl and 4-hydroxydiphenyl, which may be represented as $(\text{O}_2\text{NC}_6\text{H}_4\text{C}_6\text{H}_4\text{NO}_2)_2 (\text{C}_6\text{H}_5\text{C}_6\text{H}_4\text{OH})$. Projections of the electron density on two axial planes have been made, and the parameters of the structure measured.

The dimensions of the unit cell are

$$a = 20.06 \text{ \AA}, \quad b = 9.46 \text{ \AA}, \quad c = 11.13 \text{ \AA}, \quad \beta = 99^\circ 39'.$$

The space-group is C_m , and the unit cell contains only two of the complex groups. The hydroxydiphenyl molecules lie completely in the mirror planes (020), while the dinitrodiphenyl molecules lie across the mirror planes, and are approximately parallel to one another in the planes (20 $\bar{8}$). In this structure, therefore, the benzene rings of the hydroxydiphenyl molecules have a plane of symmetry in the plane of the rings, and the benzene rings of the dinitrodiphenyl molecules have a mirror plane passing through the terminal carbon atoms and perpendicular to the plane of the rings. The nitro groups possess a plane of symmetry through the nitrogen atom perpendicular to the plane of the group. The thermal vibrations in these crystals produce very pronounced diffuse spectra of two types, which are consistent with the presence of long, flat molecules lying in the planes (20 $\bar{8}$).

All the molecules are approximately equally spaced from one another, and the inter-molecular distances are no shorter than are normally found in aromatic nitro compounds.

The molecular dimensions determined with greatest accuracy are those between atoms which are reflected across the mirror planes, namely, the distance of 1.90 Å between the oxygen atoms of the nitro groups, and the C to C distance of 2.42 Å across the benzene rings.

INTRODUCTION

In the diphenyl series, the molecular complexes between nitro compounds on the one hand, and certain substituted hydrocarbons on the other, form a number of compounds with unusual ratios between the two components. This paper describes part of an investigation which has been made of these complexes in an attempt to obtain evidence as to the mechanism involved in their formation. An account of the chemical part of the work is to be presented elsewhere; in this paper an account is given of the crystallographic investigation. An X-ray diffraction study of similar complexes in the benzene series has already been made by Powell & Huse (1943), and Powell, Huse & Cooke (1943), and the present results are in agreement with theirs.

PREPARATION

The 4-hydroxydiphenyl used was supplied by Dr W. S. Rapsom of the University Chemistry Department; the 4:4'-dinitrodiphenyl was prepared from 4-nitrodiphenyl by the method of Gull & Turner (1929). These two components were dissolved in approximately equal proportions in warm acetone; on cooling, yellow crystals of the complex, having a melting-point of 228–229° C, were obtained. For analysis the complex was decomposed by dissolving it in as small a quantity of hot acetone as possible and adding dilute aqueous NaOH. The precipitated dinitrodiphenyl was filtered off, and the hydroxydiphenyl recovered from the filtrate by the addition of dilute HCl. The ratio of the components in the complex was found to be three parts of 4:4'-dinitrodiphenyl to one part of 4-hydroxydiphenyl, and the complex may thus be represented as $(C_6H_5C_6H_4OH)(O_2NC_6H_4C_6H_4NO_2)_3$. Crystals for X-ray analysis were grown by slowly cooling a solution of the complex in acetone containing a little excess hydroxydiphenyl, the more soluble component.

CRYSTALLOGRAPHY AND OPTICAL PROPERTIES

The crystals are canary yellow in colour, and are clear and free from flaws, so that a very complete optical examination is possible. They are monoclinic, having straight extinction on the (100) face. The faces (100) and $(\bar{1}00)$ and the prism faces {110} round the zone [001], as well as the (001) and $(00\bar{1})$ faces, often occur in the same specimen. Goniometer measurements show the monoclinic angle, β , to be 99° 39'.

The crystals are very strongly doubly refracting, and are optically positive. Under the polarizing microscope, a biaxial figure containing the acute bisectrix, which is nearly perpendicular to (100), may be obtained. The optic axial angle, $2V$, was found to be 45°.

The principal refractive indices α and β were determined by the immersion method using the Becke-line phenomenon, but solutions of high enough refractive index to determine γ in this way were not available, and it was calculated from the relation

$$\tan V = \frac{\gamma}{\alpha} \sqrt{\frac{\alpha^2 - \beta^2}{\beta^2 - \gamma^2}}.$$

The values obtained for sodium light were

$$\alpha = 1.59, \quad \beta = 1.64, \quad \gamma = 2.03.$$

The optical data are summarized in figure 1.

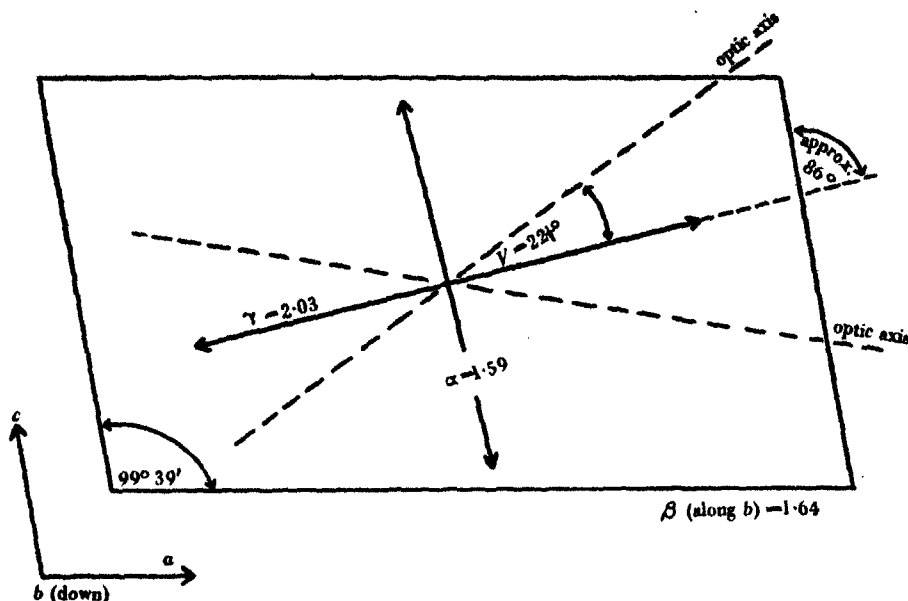


FIGURE 1. Directions of refractive indices and optic axes in the unit cell.

THE UNIT CELL AND SPACE-GROUP

Oscillation photographs were taken using filtered $\text{Cu } K_\alpha$ radiation, and the axial lengths determined from these. The dimensions of the unit cell are

$$a = 20.06 \text{ \AA}, \quad b = 9.46 \text{ \AA}, \quad c = 11.13 \text{ \AA}, \quad \beta = 99^\circ 39'.$$

The density of the crystal, determined by the method of flotation, is 1.43 g./c.c. The unit cell, therefore, contains two of the complex groups



corresponding to a true density of 1.44 g./c.c.

Reflexions of type hkl occur only with $h + k$ even. The cell, therefore, has the c -face centred. Reflexions of type $0k0$ occur only with k even, and reflexions of type $h0l$ only with h even. The space-group may, therefore, be either C_2 , $C_{2/m}$, or C_m . There are only two of the complex groups in the unit cell, and each such group contains only one molecule of hydroxydiphenyl and an odd number of molecules of dinitrodiphenyl. The general position is four-fold in the space-group C_2 and eight-fold in $C_{2/m}$, so that each molecule of hydroxydiphenyl, and at least one molecule of dinitrodiphenyl in each group, must lie in a special position. In the space-group C_2 this

special position can only be a two-fold axis. Each of the molecules concerned might have a two-fold axis, but in the case of the polar molecule of hydroxydiphenyl this can only be the long axis of the molecule. The length of this molecule, allowing for the approach of the next molecule along the two-fold axis, can hardly be less than 11.8 Å; and, since the length of the b axis of the cell, parallel to which the two-fold axes of the space-group C_2 lie, is only 9.46 Å, this would appear to exclude the possibility of this space-group. In the space-group $C_{2/m}$, since the general position is eight-fold, the odd molecules in the complex group must lie both on a two-fold axis and a mirror plane. In the case of the hydroxydiphenyl molecule this is impossible; for its only possible two-fold axis lies in its only possible mirror planes of symmetry, while in the space-group the two-fold axes are perpendicular to the mirror planes. The space-group must therefore be $C_m(C_2^2)$.

Packing considerations lead, as will be shown in the next section, to the conclusion that the hydroxydiphenyl molecule has itself to lie completely in a mirror plane; and, as there appears so far to be no case in which a benzene ring has been definitely shown to lie in a mirror plane, it was decided to test the crystal for pyro-electricity. The method used was substantially the same as that described by Orelkin & Lonsdale (1934), except that instead of an electroscope an electrometer valve was used. The sensitivity of the apparatus was tested using Rochelle salt, and, on introduction of the liquid air, the galvanometer spot was thrown off the scale. *M*-dinitrobenzene, which is known to be polar (Archer 1946), under the same conditions also gave deflexions off the scale. Using rock salt as a control, maximum deflexions of about 0.5 cm. were observed.

When a crystal of the complex of dinitrodiphenyl with hydroxydiphenyl was gripped across the b axis, maximum deflexions of the order of 1 cm. were observed on introduction of the liquid air in repeated experiments. This small deflexion, when the difficulty of holding the crystal in exactly the right direction is taken into account, may be taken as showing an absence of polarity in the b direction, which is consistent with the existence of a mirror plane perpendicular to the b axis. When the crystal was gripped across the c axis, under similar conditions, deflexions of about 10 cm. were observed, whilst across the a axis the deflexions were about 20 cm. These small, but positive, results indicate that the a and c directions are polar, and that there is no centre of symmetry; which is consistent with the space-group C_m .

PRELIMINARY ESTIMATE OF THE STRUCTURE

In the space-group C_m the general position is four-fold; it is therefore necessary so to arrange the complex groups that each has a mirror plane of symmetry, thus reducing the number of groups in the unit cell to two, and at the same time to pack the molecules into the available space. Each molecule of 4-hydroxydiphenyl and at least one of the molecules of 4:4'-dinitrodiphenyl must now lie on a mirror plane in such a way that this plane is also a symmetry plane of the molecule. In the case of the hydroxydiphenyl molecule, the OH group must lie in a mirror plane, but the

plane of the benzene rings themselves may be either in the mirror plane or at right angles to it. Three dimensional cardboard models were made of the two types of component molecule and these were packed into the unit cell with due regard to the symmetry conditions.* With normal distances of approach between adjacent benzene rings in neighbouring molecules only one arrangement is found to be possible. The hydroxydiphenyl molecules have to lie completely in the mirror planes with their greatest lengths approximately in the direction of the *c* axis; whilst the dinitrodiphenyl molecules have to lie across the mirror planes, so that the terminal carbon atoms of the benzene rings and the nitrogen atoms are in the mirror planes and the other carbon atoms and the oxygen atoms are reflected across these planes. In order to get the most economical packing, the dinitrodiphenyl molecules must lie with their greatest lengths inclined at about 11° to the *a* axis, thus interleaving with one another. The type of arrangement arrived at will easily be understood on reference to figures 5 and 7, and was indeed very close to the final structure. Confirmation of this type of arrangement is given by the refractive indices, and by the diffuse spectra due to thermal vibrations.

The direction of vibration of the light vector for the largest refractive index (γ in figure 1) is nearly parallel to the planes of all the benzene rings (see figure 5). The direction of vibration for the intermediate refractive index, β , is parallel to the planes of the benzene rings of the three dinitrodiphenyl molecules but is perpendicular to the plane of the benzene rings of the hydroxydiphenyl molecule; while the direction of vibration for the smallest refractive index, α , is parallel to the plane of the benzene rings of the one hydroxydiphenyl molecule but nearly perpendicular to the planes of the benzene rings of the three dinitrodiphenyl molecules.

Lonsdale & Smith (1941) have shown that the shape of the diffuse spectra, or optical ghosts, due to thermal vibrations in the crystal, depends on the type of structure. These ghosts are represented by diffuse distributions of the interference function around the reciprocal lattice points corresponding to the spectra. In a layer-lattice structure the distribution around the reciprocal lattice points corresponding to spectra from planes parallel to the layers of atoms will be nearly spherical, and will be cut by the sphere of reflexion in a nearly circular section. Thus, such spectra are likely to be accompanied by large and nearly circular diffuse reflexions. With chain-like structures, the reciprocal lattice points corresponding to spectra from planes perpendicular to the direction of the chains are surrounded by disc-shaped diffuse distributions of the interference function. The intersection of these by the sphere of reflexion is more or less linear and the corresponding spectra are thus accompanied by diffuse streaks. It is possible for a structure containing long, flat molecules to produce both types of ghost (Lonsdale, Robertson & Woodward 1941), and this is the case for the complex between dinitrodiphenyl and hydroxydiphenyl. The dinitrodiphenyl molecules all lie in planes approximately parallel to

* A coplanar configuration of the benzene rings in diphenyl was assumed for these models, and the dimensions were taken from the work of James, King & Horrocks (1935) on *p*-dinitrobenzene, van Niekerk (1943) on 4:4'-dinitrodiphenyl, and Robertson (1936) on resorcinol.

one another thus producing a layer-lattice type of structure, while they are also sufficiently long to have a chain-like nature.

It is found that the spectra 003 , $20\bar{3}$, $11\bar{3}$ and $1\bar{1}\bar{3}$ have very pronounced, and more or less circular, diffuse reflexions accompanying them, indicating that the dinitrodiphenyl molecules are not greatly inclined to these planes. It is also found that the spectra $10,0\bar{1}$, 802 , $91\bar{1}$ and $9\bar{1}\bar{1}$ have very pronounced diffuse streaks accompanying

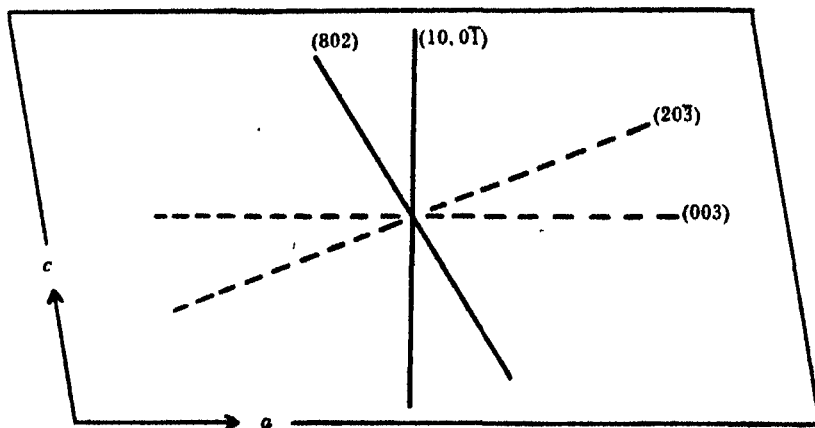


FIGURE 2. The broken lines indicate the directions in the unit cell of planes producing broad diffuse reflexions, and are thus approximately parallel to molecules lying in layers. The continuous lines indicate the directions in the unit cell of planes producing diffuse streaks, and are thus approximately perpendicular to the chain lengths of the molecules.

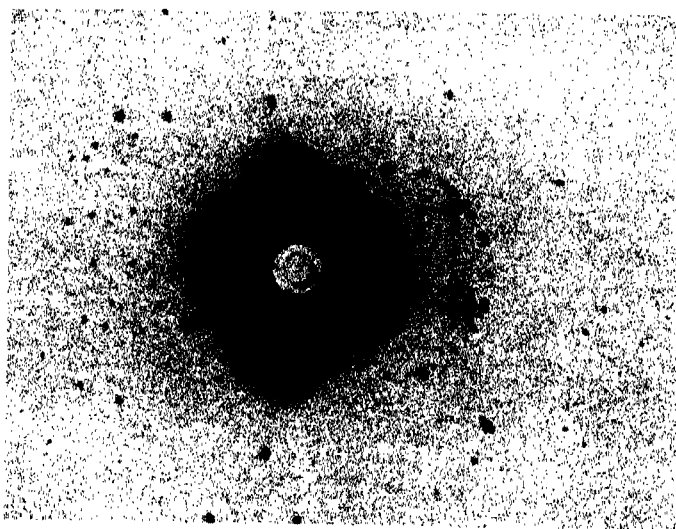


FIGURE 3. Laue photograph, taken in a cylindrical camera, with the crystal mounted with its c axis vertical, and using unfiltered copper radiation incident at 65° to a^* . The two broad diffuse areas accompany the spectra $11\bar{3}$ and $1\bar{1}\bar{3}$. The strongest streaks accompany the spectra 802 and $10,0\bar{1}$. $10,0\bar{1}$ shows streaks due to both the α - and β -radiation.

them, indicating that the lengths of the dinitrodiphenyl molecules are nearly perpendicular to these planes. A number of less pronounced ghosts corresponding to similar directions in the unit cell have been observed. The directions in the unit cell of some of the planes giving pronounced diffuse spectra are shown in figure 2 and may be compared with the actual structure as shown in figure 5. A photograph showing both types of ghost is reproduced in figure 3. It is also significant that the spectrum from the planes (020), in which lie all the hydroxydiphenyl molecules (see figure 7), also has a pronounced diffuse reflexion accompanying it, which is not circular but is considerably elongated, owing to the influence of the dinitrodiphenyl molecules.

THE FOURIER SYNTHESIS OF THE STRUCTURE

For the accurate determination of the structure the method of double Fourier series, first developed by W. L. Bragg (1929), was used. The preliminary structure promised a reasonable resolution of many of the atoms when projected on the *ac* plane, whilst a projection on the *ab* plane, although offering no actual resolution of the atoms, promised sufficient information to fix the *y* co-ordinates.

In neither case is there a centre of symmetry, and, putting the summation equations in such a form that only positive values of *h*, *k* and *l* need be considered, then, for the space-group C_m ,

$$\rho(xz) = \frac{1}{ac \sin \beta} \left[F_{(000)} + 2 \sum_0^H \sum_0^L \left\{ |F_{(h0l)}| \cos \left[2\pi \left(\frac{hx}{a} + \frac{lz}{c} \right) - \delta_{(h0l)} \right] + |F_{(\bar{h}0l)}| \cos \left[2\pi \left(-\frac{hx}{a} + \frac{lz}{c} \right) - \delta_{(\bar{h}0l)} \right] \right\} \right], \quad (1)$$

where $\rho(xz)$ is the density of the projection parallel to *b* on the *ac* plane, and

$$\rho(xy) = \frac{1}{ab} \left[F_{(000)} + 4 \sum_0^H \sum_0^K \left\{ |F_{(hko)}| \cos \left[2\pi \frac{hx}{a} - \delta_{(hko)} \right] \cos 2\pi \frac{ky}{b} \right\} \right], \quad (2)$$

where $\rho(xy)$ is the density of the projection parallel to *c* on the *ab* plane. The numerical factor in front of the summation signs must be halved if either *h* or *l* are zero in equation (1) and if either *h* or *k* are zero in equation (2). $|F_{(hol)}|$ and $|F_{(hko)}|$ are observable quantities, but the phase differences δ have to be calculated using the positions of the atoms assumed in the preliminary structure.

In the space-group C_m the co-ordinates of equivalent point positions are

$$\begin{aligned} x, y, z: & \quad x + \frac{1}{2}, y + \frac{1}{2}, z, \\ x, \bar{y}, z: & \quad x + \frac{1}{2}, \bar{y} + \frac{1}{2}, z. \end{aligned}$$

Since there are only two complex groups in the unit cell, this means that not all the atoms in a group are symmetrically independent of one another. For the purposes of structure-factor calculations, only those atoms in a complex group which are not connected by symmetry relationships need be considered.

The structure factors for atoms in general positions are given by

$$F = A + iB,$$

where

$$\left. \begin{aligned} A(h0l) &= 4\Sigma \left[f \cos 2\pi \left(\frac{hx}{a} + \frac{lz}{c} \right) \right] \\ B(h0l) &= 4\Sigma \left[f \sin 2\pi \left(\frac{hx}{a} + \frac{lz}{c} \right) \right] \end{aligned} \right\} \quad (3)$$

$$\left. \begin{aligned} A(hk0) &= 4\Sigma \left[f \cos 2\pi \left(\frac{hx}{a} \right) \cos 2\pi \left(\frac{ky}{b} \right) \right] \\ B(hk0) &= 4\Sigma \left[f \sin 2\pi \left(\frac{hx}{a} \right) \cos 2\pi \left(\frac{ky}{b} \right) \right] \end{aligned} \right\} \quad (4)$$

f is the appropriate atomic scattering factor of the atom whose co-ordinates are (x, y, z) and the sum is to be taken over all the independent atoms in the complex group. In the case of atoms lying in the mirror planes the factor of 4 in equations (3) and (4) must be halved in order to avoid counting such atoms twice.

The phase differences are calculated from

$$\delta = \tan^{-1} \frac{B}{A}. \quad (5)$$

INTENSITY MEASUREMENTS

In order to evaluate the Fourier series, it was necessary to determine the structure factors $|F(h0l)|$ and $|F(hk0)|$ for all spectra of appreciable strength. For this purpose two crystal specimens were cut to approximately cubic dimensions with a side of 0.2–0.3 mm., and were set for rotation about the b and c axes, the crystals being completely immersed in the beam of $\text{Cu } K_\alpha$ radiation. For such small crystals corrections for absorption may be nearly neglected.

As no ionization spectrometer or integrating photometer was available, the intensities of the spectra were measured by means of a photographically recording microphotometer, calibrated by means of a standard wedge. The assumption was made that the peak intensity recorded by this instrument was proportional to the integrated intensity.

In order to avoid ambiguity in indexing the spectra, since no Weissenberg camera was available at this stage, successive overlapping rotations of 15° were given. Care was taken to expose and develop each film under the same conditions, but overlapping and equivalent spectra on different films were used to standardize one film with respect to another.

An attempt was made to put the relative values thus obtained for the intensities on an approximately absolute scale by comparison with a known crystal; p -dinitrobenzene was chosen as the standard, since crystals of convenient shape are easily obtained, the absorption coefficient for this is nearly the same as that for the com-

plex, and the absolute intensities for it have already been measured by James *et al.* (1935). A photograph with a crystal of the complex rotating about the *c* axis, and set to include a representative set of spectra, was taken, one side of the film being shielded by lead. A crystal of *p*-dinitrobenzene, having approximately the same size as the complex crystal, was then set to rotate about the *b* axis; the other side of the film was then exposed under as nearly as possible the same conditions for the same length of time. This procedure was repeated using different spectra, and the films were measured up on the microphotometer. In comparing the intensities of the two sets of spectra, absorption effects may be neglected since $e^{-\mu t}$ is virtually the same for the two different crystals. The main difficulty is to make the necessary correction for the size of the two crystals. Owing to the manner in which the intensities were measured they are probably not proportional to the volumes of the crystals, as would be the case had integrated intensities been used; and since only peak intensities were measured it seems probable that an increase in cross-section of the crystal, as presented to the X-ray beam, would result mainly in an increase in the size of the reflected spots without any great increase in their peak intensity. The intensities may, therefore, be assumed to be proportional to the mean thickness, *t*, of the crystals. Neglecting absorption, the expression for *I*, the intensity as measured on the microphotometer, thus becomes

$$I \sim \left[\left(\frac{e^2}{mc^2} \right)^2 \lambda^3 \left(\frac{1 + \cos 2\theta}{2 \sin \theta} \right) \right] F_{(hkl)}^2 N^2 t.$$

The ratio $I(\text{complex})/I(p\text{-dinitrobenzene})$ may then be calculated and the intensities for the complex placed on an absolute scale. There are many possible sources of error in this method of calibration, and the *F* values of the complex so obtained must be regarded as approximate only.

There is in this case, however, a more reliable means of standardization. The *y* co-ordinates of a large number of atoms are fixed from space-group and packing considerations. Since the hydroxydiphenyl molecules lie in the mirror planes, and the dinitrodiphenyl molecules across the mirror planes, the only atoms whose co-ordinates are not fixed are the oxygens of the nitro groups and some of the carbons of the dinitrodiphenyl molecules, which are reflected, however, across the mirror planes. The dimensions of the benzene ring and the nitro group are known from previous work, and thus the *y* co-ordinates of these few atoms can also be fixed without making many assumptions. This means that the structure factors $|F(0k0)|$ can be calculated with considerable accuracy.* If the observed values for $|F(0k0)|$ are then adjusted to fit these, a correction factor to reduce all *F* values to an approximately absolute scale can be obtained.

* A rough temperature correction for the various atomic scattering factors was made and used in all subsequent structure factor calculations. For carbon the values obtained by Robertson (1935) for anthracene were used; for nitrogen and oxygen the values of James & Brindley (1931) were reduced in the same ratio as that by which Robertson's values for carbon differed from those of James & Brindley.

Agreement between the absolute F values as calculated by both methods is almost exact; this is satisfactory, but the closeness of the agreement must be considered as largely fortuitous.

EVALUATION OF THE FOURIER SERIES FOR THE PROJECTION ON THE ac PLANE

Since the calculation of structure factors according to equations (3) promised to be a very laborious process, for the first approximation the following assumptions were made:

(a) That, if the OH group be left out of account, the structure has a centre of symmetry at the origin. This assumption is consistent with a reasonable packing, and the relatively small polarity observed in the a and c directions suggests that it may not be far from the truth.

(b) That the dinitrodiphenyl molecules lie parallel to one another and are evenly spaced with a separation of $\frac{1}{2}c$ in the c direction. This is borne out by the strength of the 003, 20 $\bar{3}$, and particularly the 20 $\bar{6}$ spectra.

(c) That the dinitrodiphenyl molecules are completely planar, although to get the best packing it is necessary to tilt the nitro groups slightly away from the plane of the benzene rings. The structure-factor calculations were thus considerably simplified, while the agreement between $|F(\text{obs.})|$ and $|F(\text{calc.})|$ was on the whole quite good.

The phase relationships for the various spectra were calculated as follows: since a centre of symmetry had been assumed for the whole structure, excepting the OH group, the value of B was given by the position of this group alone. The magnitude of A was taken from

$$|A| = \sqrt{(F_{\text{obs.}}^2 - B^2)},$$

while the *sign* of A was taken as being that of its calculated value. The phases were then calculated according to the expression (5).

The series (1) was now evaluated using the method of Lipson & Beevers (1936), and the projection so obtained was fairly clean with little false detail, except around the position of the hydroxydiphenyl molecule. From this projection, improved values of the x and z co-ordinates were estimated. For most atoms there was little change, but a tilt of the nitro groups away from the plane of their benzene rings had appeared. Except for the OH group, there was no departure from a centre of symmetry at the origin for the structure.

The structure factors were recalculated still assuming this approximation to a centre of symmetry, but without other simplification. The values of A and B used in determining the phase relationships for the summation were obtained in the same way as before, and the series was again evaluated. The projection so obtained was much cleaner, and showed better resolution than the first, but there was little appreciable change in the positions of the peaks, even those of the nitro groups.

The observed and calculated structure factors are shown in table 1.

The agreement on the whole is excellent and generally it is only for spectra with $\sin \theta$ greater than 0.5 that any serious discrepancies are found. As no *differential* corrections for the temperature motion were made for the different spectra, one would expect the observed values for those spectra, particularly the higher orders, which are most affected by thermal vibration, to be somewhat lower than the calculated values. Such spectra will be produced by those planes which most nearly contain the dinitrodiphenyl molecules and so have the largest vibrations approximately perpendicular to them. These spectra are mainly 003, 006 and $20\bar{3}$, $40\bar{6}$ and $20\bar{6}$, $40\bar{1}\bar{2}$. It is found that the calculated values for 006, $40\bar{6}$ and $40\bar{1}\bar{2}$ are far too high.

The contour diagram drawn from the final values of $\rho(xz)$ for the projection on the ac face viewed parallel to b is shown in figure 4. This contour diagram may be interpreted with the aid of the key diagram shown in figure 5.

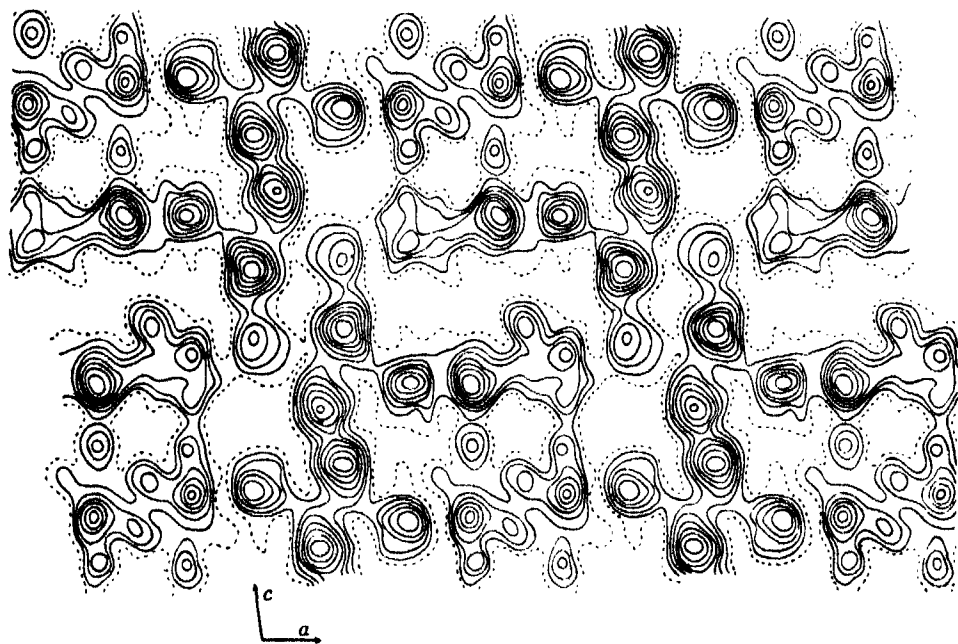


FIGURE 4. Fourier projection of the structure in the direction of the b axis on to the ac plane. Contours at intervals of approximately 2 electrons per \AA^2 . The dotted line is the 2-electron contour.

In all projections where there is no centre of symmetry the observed coefficients $|F|$ must be combined with phase differences δ that have to be calculated, according to equation (5), from assumed data. In the present instance, the assumptions made have led to no inconsistencies, and there is no reason to believe that they can represent any marked deviation from the truth. Only the position of the OH group has been used to determine B , since a centre of symmetry was assumed for the rest of the structure. B^2 is generally very small compared with $F^2(\text{obs.})$, and, with the

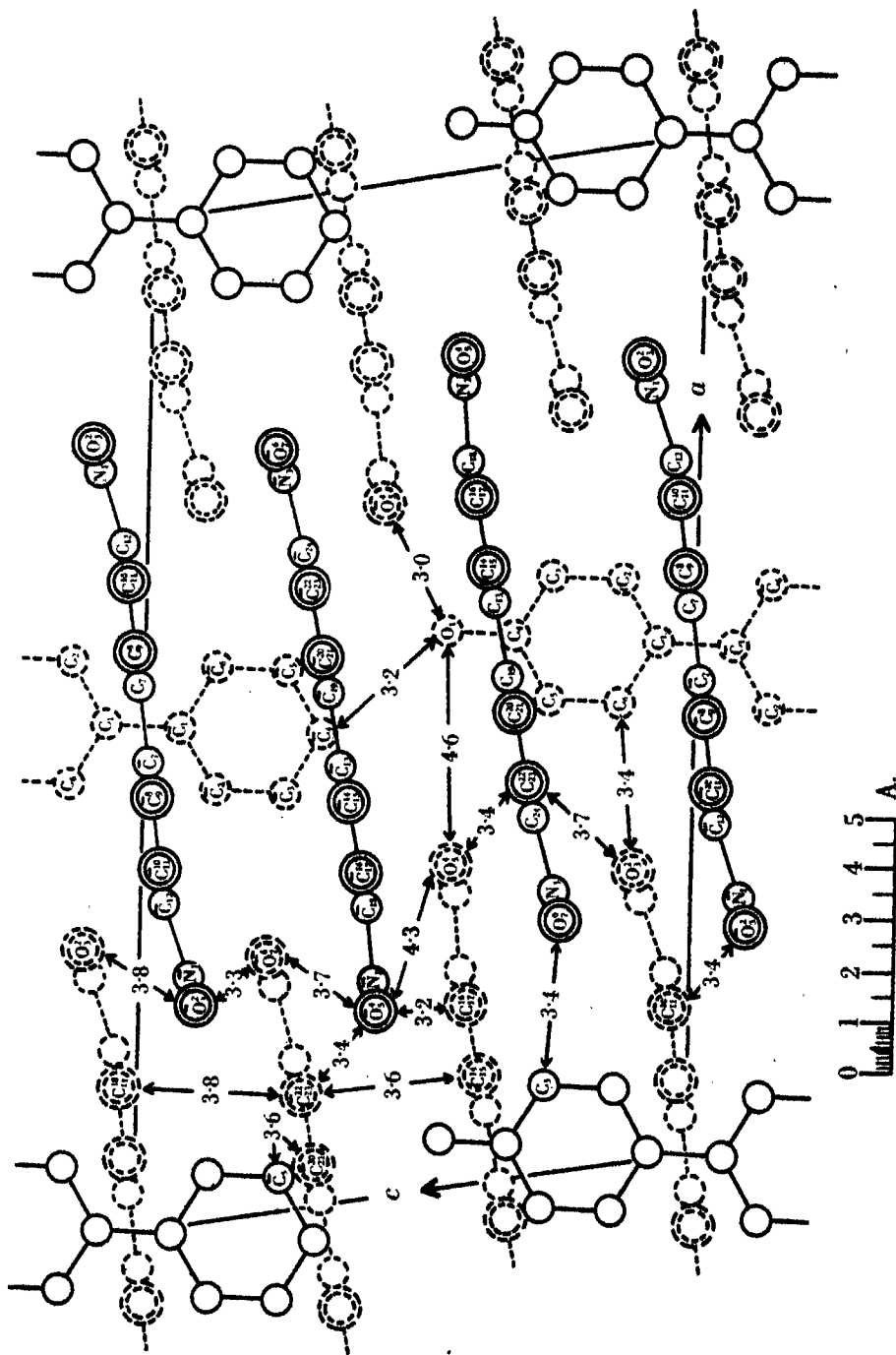


FIGURE 5. Projection of the structure in the direction of the *b* axis on to the *ac* plane. The molecules denoted by broken lines lie $\frac{1}{2}$ above or below those denoted by unbroken lines. Double circles indicate two atoms superimposed upon one another by reflexion across the mirror planes.

TABLE 1. OBSERVED AND CALCULATED STRUCTURE FACTORS

<i>hkl</i>	$ F(\text{obs.}) $	$ F(\text{calc.}) $	<i>hkl</i>	$ F(\text{obs.}) $	$ F(\text{calc.}) $
200	60	60	13,30	13	7
400	113	103	14,40	33	15
800	97	65	14,60	13	4
800	54	33	15,30	21	15
10,00	35	27	15,50	17	13
12,00	8	7	16,20	10	7
14,00	27	16	16,40	17	13
			16,60	10	10
020	300	318	17,10	12	8
040	35	42	19,10	8	12
060	80	76			
080	149	134	001	14	25
0,10,0	66	56	002	24	28
			003	213	181
110	68	55	004	20	10
130	139	112	005	47	29
150	81	51	006	36	60
170	27	26	007	32	14
190	30	4	00,12	24	5
220	27	12			
240	99	85	201	42	18
260	68	55	202	63	25
280	23	4	203	22	33
310	89	52	204	17	17
330	52	41	205	39	18
350	58	55	206	18	15
370	47	38	208	19	7
390	20	5	209	37	54
420	130	120	20,10	19	4
440	102	74	201	35	13
460	39	12	202	41	26
4,10,0	15	29	203	180	151
510	112	86	204	37	59
530	95	86	205	53	64
550	100	82	206	107	241
590	16	18	207	14	47
620	41	22	401	26	14
640	29	19	402	22	28
710	115	80	403	90	126
730	9	9	404	28	28
750	27	25	405	34	59
820	27	16	406	14	28
840	11	6	408	13	14
910	43	43	401	91	79
930	21	20	402	42	33
950	18	19	403	24	24
990	12	12	404	52	33
10,20	19	18	405	29	30
10,40	10	12	406	25	104
10,60	12	15	408	12	12
11,10	11	13	409	17	23
11,30	12	13	40,12	—	55
13,10	10	7			

Table 1 (*continued*)

hkl	$ F(\text{obs.}) $	$ F(\text{calc.}) $	hkl	$ F(\text{obs.}) $	$ F(\text{calc.}) $
601	12	11	10,0 $\bar{1}$	190	123
602	114	91	10,0 $\bar{2}$	85	106
603	56	73	10,0 $\bar{3}$	15	13
604	18	32	10,0 $\bar{4}$	26	14
607	21	11	12,01	19	15
60 $\bar{1}$	113	97	12,0 $\bar{5}$	16	42
60 $\bar{2}$	59	44	12,07	26	12
60 $\bar{3}$	14	43	12,08	28	29
60 $\bar{6}$	28	46	12,0 $\bar{1}$	59	21
607	14	30	12,0 $\bar{2}$	19	7
60 $\bar{8}$	28	30	12,0 $\bar{4}$	15	44
801	74	58	12,07	42	67
802	148	105	14,01	40	14
803	60	79	14,02	39	52
804	27	42	14,0 $\bar{1}$	81	68
805	75	58	14,0 $\bar{2}$	37	26
806	21	27	16,01	28	11
807	30	28	16,02	17	7
808	17	2	16,04	19	10
80 $\bar{1}$	50	35	16,0 $\bar{1}$	15	3
80 $\bar{2}$	52	48	16,0 $\bar{2}$	47	33
80 $\bar{3}$	90	83	16,07	53	12
80 $\bar{4}$	55	65	16,0 $\bar{8}$	20	3
807	17	52	16,0,1 $\bar{0}$	17	2
80 $\bar{8}$	32	4	18,01	23	17
80,1 $\bar{0}$	37	4	18,02	18	26
10,01	42	26	18,04	26	28
10,02	47	36	18,0 $\bar{2}$	34	20
10,05	20	4	18,0,1 $\bar{0}$	20	16
10,06	21	20	20,0 $\bar{2}$	18	19
10,07	16	18	20,0 $\bar{5}$	23	13

method of calculation used for the phase differences, the *magnitude* of A thus depends mainly on observation, the *sign* only being determined from the assumed positions of the atoms. It is significant that the benzene rings of the hydroxydiphenyl molecule appear in the projection as clearly as can be expected, although their position rarely affected the *sign* of A which was determined almost entirely by the dinitrodiphenyl molecules. It is also satisfactory that the tilt of the nitro groups to the plane of the molecules has emerged in the course of the summations and was not assumed initially. Probably the most convincing confirmation of the proposed structure, however, is the excellent agreement between $|F(\text{obs.})|$ and $|F(\text{calc.})|$.

EVALUATION OF THE FOURIER SERIES FOR THE PROJECTION ON THE ab PLANE

From the previous projection the x co-ordinates of the atoms can be estimated while, as has already been shown, the y co-ordinates can be fixed from the preliminary structure with some accuracy.

The structure factors $F(hk0)$ were calculated according to the expression (4), assuming as before a centre of symmetry for the structure, with the exception of the

OH group. The values of $|F(\text{obs.})|$ and $|F(\text{calc.})|$ are included in table 1. The agreement is excellent. The phase relationships were calculated in the same way as for the *ac* projection; thus the value of *B* was given by the OH group alone, and the sign of *A* was taken as being that of its calculated value while its magnitude was calculated from

$$|A| = \sqrt{(F_{(\text{obs.})}^2 - B^2)}.$$

The series (2) was then evaluated using, as before, the method of Lipson & Beevers (1936). The contour diagram and key diagram corresponding to values of $\rho(xy)$ are given in figures 6 and 7.

ESTIMATION OF PARAMETERS

From the *ac* projection, the *x* and *z* parameters were estimated as follows:

(a) *The dinitrodiphenyl molecules.* The displacement of the position of $C_{8,9}$ from the centre of the peak in which they are unresolved from C_7 was assumed to be the same as the displacement of $C_{10,11}$ from the centre of the peak in which they are unresolved from C_{12} . If the side of the benzene ring is 1.40 Å this displacement is approximately 0.13 Å, and, assuming the benzene ring to be a regular hexagon, the positions of all the carbon atoms in this ring can be fixed. An estimate was made of the sort of displacement to be expected, taking into account the height of the peaks produced by the separate atoms, 0.7 Å apart, and the approximate resolving power shown in the projection, and agreement was good. A similar displacement was calculated for the unresolved $O_{2,3}$ - N_1 peak and the position of the two oxygen atoms fixed at about 0.20 Å from the centre of the peak, and directly away from C_{12} . The position of the nitrogen atom was then fixed by assuming a C-N distance of 1.53 Å, as found by James *et al.* (1935), and placing the nitrogen along the line joining C_{12} to $O_{2,3}$. This procedure was repeated for the other benzene rings and nitro groups. The benzene rings in each dinitrodiphenyl molecule were assumed to be coplanar, and the slight deviations from this shown by the projection were neglected.

(b) *The hydroxydiphenyl molecule.* Here, two benzene rings, regular hexagons of side 1.40 Å, were adjusted so as to give the best fit with the resolved peaks C_1 , C_2 , C_6 and \bar{C}_1 , \bar{C}_2 , \bar{C}_6 , while preserving a linear arrangement of the atoms C_4 , C_1 , \bar{C}_1 , \bar{C}_4 . The OH peak is resolved.

The estimation of the *y* parameters from the *ab* projection presented no difficulties. A list of parameters is given in table 2.

DESCRIPTION OF THE STRUCTURE

The packing of the molecules in this structure is best illustrated by figure 7. It may be described as a face-centred arrangement of dinitrodiphenyl molecules, all lying across the mirror planes (020) and nearly in the planes (20 $\bar{6}$), thus interleaving with one another (as shown in figure 5); the long holes thus left in the structure are filled by the hydroxydiphenyl molecules lying nearly end to end in the mirror planes

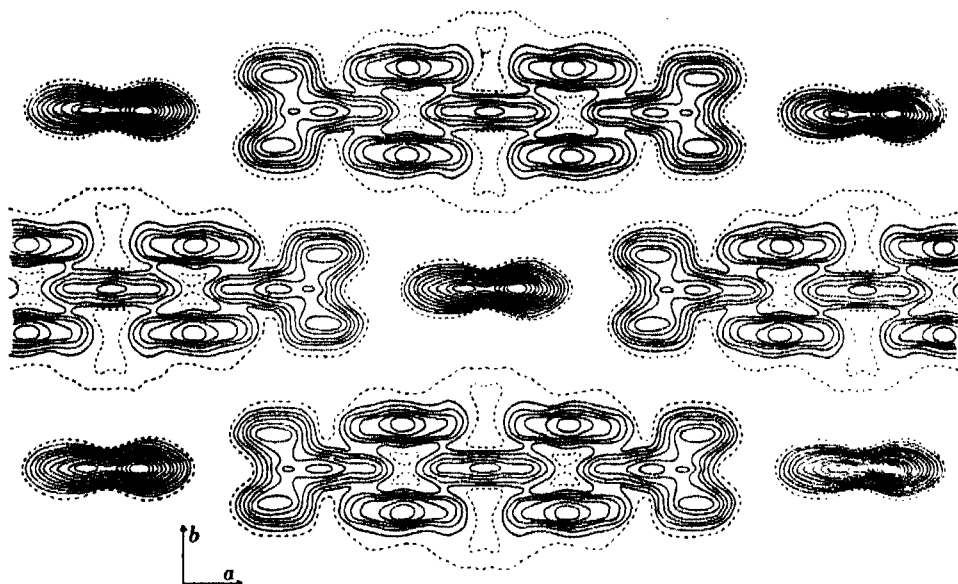


FIGURE 6. Fourier projection of the structure in the direction of the c axis on to the ab plane. Contours at intervals of approximately 2 electrons per \AA^2 . The dotted line is the 2-electron contour.

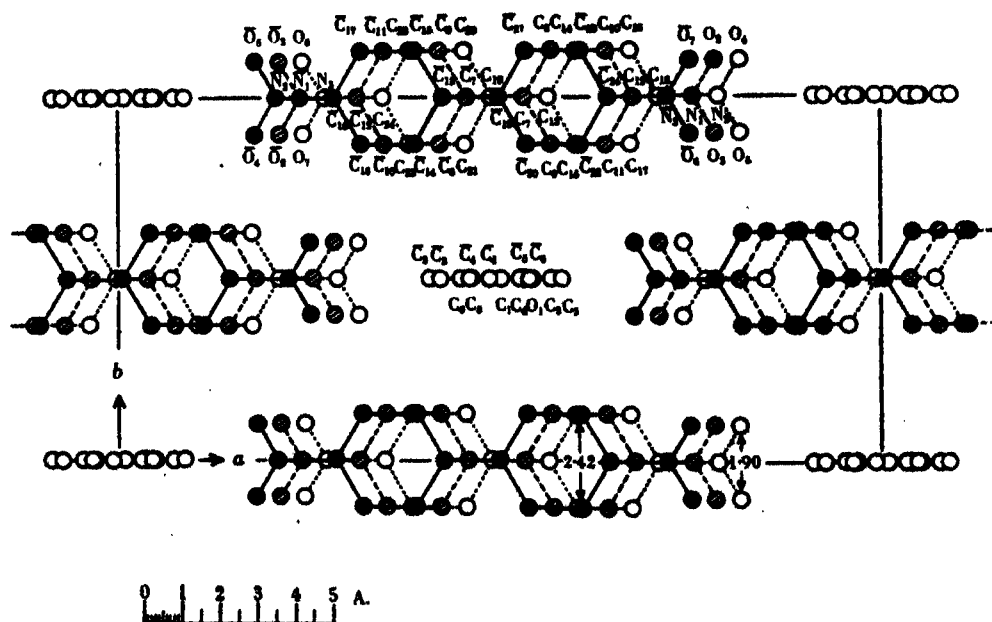


FIGURE 7. Projection of the structure in the direction of the c axis on to the ab plane. The dinitrodiphenyl molecules denoted by black circles are at a height $2c/3$, those denoted by plain circles at a height $c/3$, and those denoted by shaded circles at a height $0c$.

TABLE 2. PARAMETERS OF THE ATOMS EXPRESSED AS FRACTIONS
OF THE CORRESPONDING LATTICE TRANSLATIONS

The atoms denoted by bars in the figures 5 and 7 are to be taken as derived from their corresponding atoms listed below by the operation of a centre of symmetry at the origin. The actual deviation from this was very small and has been neglected.

atom	x/a	y/b	z/c	atom	x/a	y/b	z/c
C ₁	0.007	0	0.068	C ₁₉	-0.006	0.500	0.337
C ₂	0.073	0	0.131	C ₂₀	-0.041	0.628	0.324
C ₃	0.088	0	0.258	C ₂₁	-0.041	0.372	0.324
C ₄	0.034	0	0.323	C ₂₂	-0.112	0.628	0.298
C ₅	-0.032	0	0.260	C ₂₃	-0.112	0.372	0.298
C ₆	-0.047	0	0.132	C ₂₄	-0.147	0.500	0.286
C ₇	0.038	0.500	0.010				
C ₈	0.073	0.628	0.018	O ₁	0.048	0	0.445
C ₉	0.073	0.372	0.018	O ₂	0.285	0.600	0.103
C ₁₀	0.144	0.628	0.036	O ₃	0.285	0.400	0.103
C ₁₁	0.144	0.372	0.036	O ₄	0.317	0.600	0.430
C ₁₂	0.179	0.500	0.045	O ₅	0.317	0.400	0.430
C ₁₃	0.069	0.500	0.363	O ₆	-0.253	0.600	0.235
C ₁₄	0.105	0.628	0.376	O ₇	-0.253	0.400	0.235
C ₁₅	0.105	0.372	0.376				
C ₁₆	0.176	0.628	0.402	N ₁	0.256	0.500	0.088
C ₁₇	0.176	0.372	0.402	N ₂	0.288	0.500	0.427
C ₁₈	0.211	0.500	0.414	N ₃	-0.224	0.500	0.247

with their lengths nearly normal to the planes containing the dinitrodiphenyl molecules. Each hydroxydiphenyl molecule is thus surrounded by twelve dinitrodiphenyl molecules all approximately equally spaced from it. In these circumstances it is impossible to choose any particular group of individual molecules as representing the 'complex molecule', the discrete existence of which must be doubted. With a view to illustrating the molecular dimensions of the individual components in the structure, however, a group of three dinitrodiphenyl molecules and one hydroxydiphenyl molecule, which may be taken as representing the repeating unit of structure, is shown in figure 8.

An unusual and striking feature of this structure is that certain groupings must be perfectly symmetrical. Thus the benzene rings of the hydroxydiphenyl molecules have a plane of symmetry in the plane of the rings, and the benzene rings of the dinitrodiphenyl molecules have a mirror plane passing through the terminal carbon atoms and perpendicular to the plane of the rings. It, therefore, seems reasonable to infer that *each* benzene ring possesses both types of mirror plane. The nitro groups must also possess a plane of symmetry, passing through the nitrogen atom, and across which the two oxygen atoms are reflected; this suggests that the unequal N-O distances observed by James *et al.* (1935) and by van Niekerk (1943) in *p*-dinitrobenzene and 4:4'-dinitrodiphenyl respectively, may not have been a real effect.

The molecular dimensions of greatest interest are those of the nitro group. A distance of 1.90 Å is found between the two oxygen atoms which are reflected across

the mirror plane; and, since the distance between the carbon atoms of the benzene ring, similarly reflected across the mirror plane, is 2.42 Å (which agrees well with the corresponding dimension, 2.425 Å, for a regular hexagon of side 1.40 Å), this is probably a fairly accurate determination. This distance of 1.90 Å is somewhat less than the O-O distances in the nitro group found by other workers, namely, 2.14 Å in *p*-dinitrobenzene (James *et al.* 1935), 2.00 Å in 4:4'-dinitrodiphenyl (van Niekerk 1943), and 2.17 Å in *m*-dinitrobenzene (Archer 1946). Unfortunately, the positions of the nitrogen atoms cannot be fixed with accuracy, but by assuming a C-N distance of 1.53 Å a value of 1.10 Å is obtained for the N-O distance in the nitro groups. This may be compared with the values 1.25 and 1.10 Å obtained by James *et al.* (1935), 1.21 and 1.14 Å by van Niekerk (1943), and 1.20 Å by Archer (1946). The tilt of the nitro groups away from the plane of the benzene rings, varying from 5 to 11°, seems to be due to packing considerations, and is such as to give a reasonable balance between the O-O and O-CH distances between interleaving molecules, without making any of these distances abnormally short.

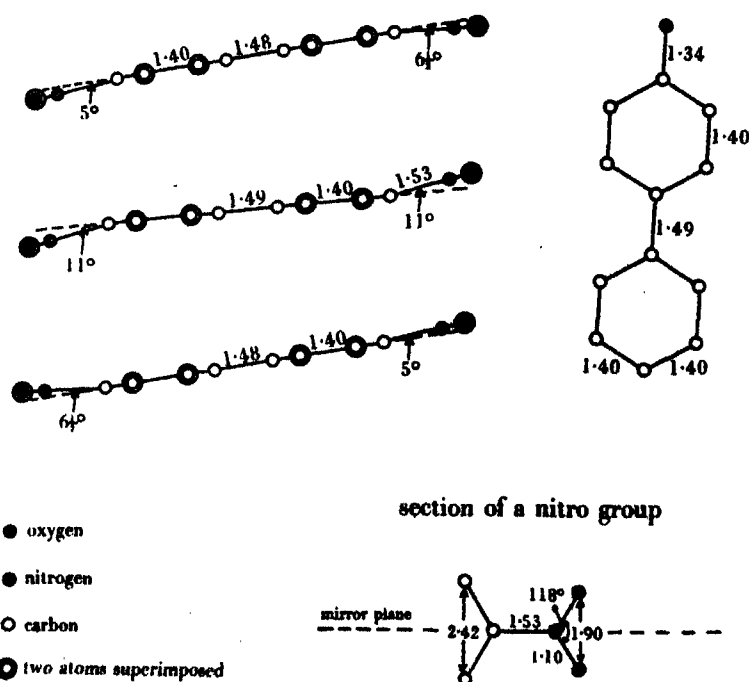


FIGURE 8. A repeating unit of structure projected on to one of the mirror planes.

Also of interest are the C-C distances, between the benzene rings in the diphenyl part of the molecules, of 1.48–1.49 Å; these may be compared with the values of 1.48 Å obtained by Dhar (1932), and 1.42 Å by van Niekerk (1943). The present values are not very accurate but suggest that the diphenyl link is nearer 1.5 than 1.4 Å.

The intermolecular distances, the most important of which are shown in figure 5, are not abnormal when compared with those so far encountered in other aromatic nitro compounds (see table 3).

TABLE 3. DISTANCES OF CLOSEST INTERMOLECULAR APPROACH
OBSERVED IN VARIOUS NITRO COMPOUNDS

compound	O-O (between nitro groups)	O-CH (between nitro groups and benzene rings)	CH-CH (between benzene rings)
complex of 4:4'-dinitrodiphenyl with	A	A	A
4-hydroxydiphenyl	3.3	3.2	3.6
<i>p</i> -dinitrobenzene	3.1 ₈	3.3 ₀	3.6 ₈
<i>m</i> -dinitrobenzene	3.2 ₀	2.9 ₅	3.6 ₀
4:4'-dinitrodiphenyl	3.4	2.9	3.6
complex of <i>s</i> -trinitrobenzene with			
<i>p</i> -iodoaniline	3.2 ₆	3.3 ₆	3.5

The distances between oxygens in neighbouring nitro groups are 3.8, 3.3, 3.7, 4.3 Å between O₃ and O₃, O₃ and O₆, O₆ and O₅, O₅ and O₅, respectively. Some of the O-CH links between neighbouring dinitrodiphenyl molecules are 3.2, 3.4, 3.4, 3.6, 3.7, 3.8 Å between O₅ and C₁₇, O₅ and C₂₂, O₃ and C₁₁, O₆ and C₁₇, O₃ and C₂₂, O₆ and C₁₁, respectively; while some similar links between nitro groups and hydroxydiphenyl molecules are 3.4, 3.4, 3.8, 3.6, 3.7, 3.8 Å between O₆ and C₃, O₂ and C₆, O₃ and C₅, O₆ and C₂, O₄ and C₅, O₄ and C₄, respectively. The closest C-C distances are quite normal, namely, 3.6, 3.6, 3.8 Å between C₅ and C₂₀, C₁₅ and C₂₂, C₁₁ and C₂₂, respectively. The closest intermolecular approach of all is that between the OH group and one of the nitro groups, i.e. 3.0 Å between O₁ and O₅.

THE BONDING

In this structure all the molecules are approximately equally spaced from one another so that it is not possible to choose any particular group of individual molecules as representing the complex group, and none of the intermolecular distances are incompatible with ordinary van der Waal's forces. These results are thus in agreement with those obtained by Powell, Huse & Cooke (1943) for the complex between *s*-trinitrobenzene and *p*-iodoaniline.

The closest intermolecular approaches are those between oxygen atoms of nitro groups and carbon atoms of benzene rings, but it is unlikely that these would provide a mechanism for the formation of the complex, since the approach of the nitro groups to the benzene rings of the hydroxydiphenyl molecule is no closer than their approach to the benzene rings of the dinitrodiphenyl molecules (in fact, the latter values are, on the whole, smaller); moreover, these distances are of the same order as those observed in crystals of ordinary aromatic nitro compounds such as *p*-dinitrobenzene

(James *et al.* 1935), *m*-dinitrobenzene (Archer 1946), and 4:4'-dinitrodiphenyl (van Niekerk 1943).

There is a possibility of weak hydrogen bonding between the OH group and the oxygen atoms of one of the nitro groups. The OH-O distance is 3.0 Å, and corresponding NH-O distances in urea have been attributed to hydrogen bonding. This, however, cannot provide a bonding mechanism in this complex where the ratio of dinitrodiphenyl to hydroxydiphenyl is 3:1; moreover, other complexes have been observed with 4:4'-dinitrodiphenyl in which there is no hydrogen available for hydrogen bonding.

The density of this complex, 1.43 g./c.c., is no higher than for ordinary nitro compounds such as *p*-dinitrobenzene (1.64), *m*-dinitrobenzene (1.57) and 4:4'-dinitrodiphenyl (1.45). In addition, the very pronounced thermal vibrations in the crystal of this complex, as shown by the strong diffuse spectra, are an indication of weak bonding between the molecules.

A discussion of these results in the light of the various theories that have been advanced as to the bonding in complexes of this type will be given in the chemical paper which is to be presented elsewhere.

Very similar structures have been found, though not worked out in detail, for the complexes of 4:4'-dinitrodiphenyl with 4:4'-dihydroxydiphenyl, benzidine, N,N:N',N'-tetramethylbenzidine, 4-iodo- and 4-bromodiphenyl. In all these complexes the ratio of the components seems to depend mainly on packing considerations, and is determined largely by the length of the molecule which is combined with the dinitrodiphenyl. In these complexes, also, there is no necessity to assume abnormal intermolecular distances in order to get reasonable packing. It is hoped to publish an account of the preliminary structures of these complexes soon.

In conclusion, I wish to thank Professor R. W. James for his constant interest and aid throughout the course of this work, and Dr W. S. Rapson for having first suggested the problem and for his help on the chemical side. I am in debt to Professor F. Walker for the use of apparatus in the Geology Department, and to Mrs F. C. M. Mathias of the same department for advice and help in the measurement of the optical properties of the crystals. I am grateful to Miss E. M. Archer for the use of unpublished data on the structure of *m*-dinitrobenzene. To the Beit Railway Trust Board of Southern Rhodesia I am indebted for a research fellowship.

REFERENCES

- Archer 1946 *Proc. Roy. Soc. A*, **188**, 51.
Bragg, W. L. 1929 *Proc. Roy. Soc. A*, **123**, 537.
Dhar 1932 *Indian J. Phys.* **7**, 43.
Gull & Turner 1929 *J. Chem. Soc.* p. 494.
James & Brindley 1931 *Phil. Mag.* **12**, 81.
James, King & Horrocks 1935 *Proc. Roy. Soc. A*, **153**, 225.
Lipson & Beevers 1936 *Proc. Phys. Soc.* **48**, 772.

- Lonsdale, Robertson & Woodward 1941 *Proc. Roy. Soc. A*, **178**, 43.
Lonsdale & Smith 1941 *Proc. Roy. Soc. A*, **179**, 8.
Orelkin & Lonsdale 1934 *Proc. Roy. Soc. A*, **144**, 633.
Powell, Huse & Cooke 1943 *J. Chem. Soc.* (1), p. 153.
Powell & Huse 1943 *J. Chem. Soc.* (2), p. 435.
Robertson 1935 *Proc. Roy. Soc. A*, **150**, 106.
Robertson 1936 *Proc. Roy. Soc. A*, **157**, 79.
van Niekerk 1943 *Proc. Roy. Soc. A*, **181**, 314.
-

The crystal structure of meta-dinitrobenzene

By E. M. ARCHER, M.Sc., *Porter Scholar of the University of Cape Town*

(Communicated by Sir Laurence Bragg, F.R.S.—Received 22 January 1946)

The crystal of meta-dinitrobenzene, $C_6H_4(NO_2)_2$, is orthorhombic pyramidal. The dimensions of the unit cell, which contains 4 molecules, are $a = 13.3$ Å, $b = 14.1$ Å, $c = 3.80$ Å and the space group is *Pbn*.

The method of double Fourier series has been applied, and a projection of the electron density on the *ab* plane has been made.

The plane of the benzene ring of the molecule is inclined at an angle of 20° to the *b*-axis, and is parallel to the *a*-axis. Within the limits of experiment, the benzene ring is a regular hexagon of side 1.41 Å. The C—N links do not lie in the plane of the ring, but make an angle of 15° with it. The C—N distance is 1.54 Å, the N—O distances have been assumed the same, 1.20 Å, and the O—O distance is 2.17 Å.

The closest approach of O to CH in adjacent molecules is about 3.0 Å, that between O and O about 3.2 Å, and that between CH and CH is 3.8 Å. A discussion of the packing of the molecules in the structure is given.

INTRODUCTION

Meta-dinitrobenzene forms pale straw-coloured needle-shaped crystals belonging to the orthorhombic system, which melt at 91° C. Von Groth gives the symmetry as orthorhombic bipyramidal, the axial ratios, as determined by Steinmetz (1915), being $a:b:c = 0.9435:1:0.5434$, with the *c*-axis in the direction of elongation of the needles. Good, clear crystals may be obtained by slowly cooling a solution in a mixture of alcohol and acetone. Faces giving goniometer reflexions develop around the [001] zone, the common ones being (100) and the prism faces (120) and ($\bar{1}20$). The unit cell, as determined by means of X-rays, has the dimensions $a = 13.3$ Å, $b = 14.1$ Å and $c = 3.80$ Å, which correspond to the axial ratios

$$a:b:c = 0.943:1:0.270.$$

The indices given in this paper refer to these axial ratios and not to those of Steinmetz. Previous attempts to determine the structure of the crystal by means of X-rays have been made by Hertel (1930), Hendricks & Hibbert (1931), and Banerjee & Ganguly (1940). All these workers have assumed the symmetry of the crystal to

be orthorhombic bipyramidal, and when the very small value of the c -spacing is taken into account, this assumption allows only two possible types of structure, both of which are physically improbable and neither of which is consistent with the observed intensities. This was recognized by Hendricks & Hibbert and, shortly after the publication of their paper, it was discovered by Hendricks* that the c -axis of the crystal was in fact strongly polar when tested for piezoelectricity. This polarity was confirmed by James (unpublished), and has again been tested in the present work. The crystal when cooled in liquid air, and tested for electrification in the manner suggested by Lonsdale & Orelkin (1934), except that instead of an electroscope an electrometer valve was used, showed a strong effect in the direction of the c -axis, but none in the directions of the a - and b -axes. The symmetry of the crystal is therefore not orthorhombic bipyramidal, but orthorhombic pyramidal and, since there is no longer a mirror plane perpendicular to the c -axis, much greater freedom in the choice of types of structure is possible.

Banerjee & Bhattacharjya (1938), from observations of the magnetic anisotropy of the crystals, have concluded that the planes of the benzene rings are inclined at an angle of $22^\circ 20'$ to the (001) planes. This result, in conjunction with the X-ray data, is inconsistent with the existence of a mirror plane perpendicular to the c -axis, and confirms the symmetry of the crystal as being orthorhombic pyramidal. It is also entirely inconsistent with the structure proposed by Banerjee & Ganguly (1940) as a result of X-ray analysis.

THE UNIT CELL AND SPACE GROUP

The density of the crystal is 1.570 g./c.c., which corresponds to 4 molecules of $C_6H_4(NO_2)_2$ in a unit cell having the dimensions given above.

Oscillation photographs were taken about all three axes with $Cu K_\alpha$ radiation. General reflexions hkl of all types occur, showing that the unit cell is primitive. Reflexions $0kl$ occur only with $k = 2n$, and $h0l$ only with $h + l = 2n$. There are no restrictions for the $hk0$ spectra. The possible space groups are $Pbn(C_{2v}^2)$ corresponding to a hemihedral point group, or $Pbnm(D_{2h}^{10}$ or $V_h^{10})$ corresponding to a holohedral point group. The second of these possibilities is ruled out by the observation that the c -axis is polar. The space group is therefore Pbn .

PRELIMINARY ESTIMATE OF THE STRUCTURE

There are 4 molecules in the unit cell, which is the number corresponding to a general position in the space group Pbn . Nothing in the structure is therefore fixed by symmetry. Had the space group been $Pbnm$, the assumption made by previous workers, the general positions would have been eightfold. In order to reduce the number of molecules in the unit cell to 4, it would therefore have been necessary to suppose the mirror planes possessed by this group to be also planes of symmetry of the molecule. The small c -spacing, 3.8 Å, which is about the known

* Private communication to Professor R. W. James.

thickness of the benzene ring, would make it necessary to suppose the benzene rings themselves to lie on mirror planes. The nitro groups would then have to lie either with all their atoms on the mirror planes also, or with the nitrogen atoms on the mirror planes and the oxygen atoms symmetrically disposed on either side of them. The first of these possibilities is negated at once by the observed intensities, for it would give a very strong 002 spectrum, whereas 002 is, in fact, very weak. The second type of arrangement would reduce the intensity of the 002 spectrum but would bring the oxygen atoms in adjacent molecules improbably close to one another. The distance obtained by Hendricks & Hibbert is 1.91 Å, whereas distances of over 3 Å have been found in other aromatic compounds, and distances of 2.7 Å in inorganic compounds. The structure proposed by Banerjee & Ganguly is also of the same type and is subject to the same objections as that of Hendricks & Hibbert. The latter workers were alive to these difficulties, and, because of them, Hendricks suggested that the crystal was polar, as has already been stated in the introduction.

If the space group is assumed to be *Pbn*, it is no longer necessary to suppose the benzene rings to lie on mirror planes. Nevertheless, there is a good deal of evidence for the view that the planes are not greatly inclined to the plane (001). The three principal refractive indices of the crystal, determined by James & Horrocks, and not hitherto published, are, for sodium light, $\alpha = 1.48$, $\beta = 1.68$, $\gamma = 1.71$, corresponding to vibrations parallel to *c*, *b* and *a* respectively. The small refractive index for vibrations parallel to *c* indicates that the *c*-axis is nearly perpendicular to the planes of the benzene rings, while the relative magnitudes of β and γ indicate that the planes of the rings are nearly parallel to *a* and inclined at a not very large angle to *b*.

An examination of the diffuse reflexions, the optical ghosts due to thermal vibrations, accompanying some of the spectra, points in the same direction. Laue photographs taken with unfiltered copper radiation when the *c*-axis is vertical show a remarkable set of diffuse reflexions in positions nearly corresponding to the first layer lines of a rotation photograph about the same axis. The ghosts accompanying the spectra 021 and $0\bar{2}1$ are particularly strong and diffuse and are of the layer lattice type (Lonsdale 1942), and this suggests that the planes of the benzene rings lie not far from the (021) and ($0\bar{2}1$) planes.

In making use of these data to estimate a preliminary structure of the crystal it was assumed that the benzene ring was a regular hexagon of side 1.4 Å, and the size and shape assumed for the nitro groups were based on the results of James, King & Horrocks (1935) for para-dinitrobenzene and of van Niekerk (1943) for 4:4'-dinitrodiphenyl. Three-dimensional cardboard models were constructed to scale and packed into the unit cell in such a way as to fulfil the symmetry conditions, and to conform to the arrangements suggested by the refractive indices and the thermal ghosts. In this way it was possible to reduce the likely arrangements to not too great a number, which could be tested by comparison between observed and calculated intensities. The arrangements which agreed best in this respect were then used as bases for Fourier syntheses.

RELATIVE INTENSITY MEASUREMENTS

All the observations were photographic, and were made with a cylindrical oscillation camera with $\text{Cu } K_\alpha$ radiation filtered through nickel foil. Neither an ionization spectrometer nor an integrating photometer was available, so that it was not possible to obtain absolute measurements and a set of relative measurements had to suffice. The film was passed in front of a photographically recording microphotometer, which had been previously calibrated by Dr J. N. van Niekerk by means of a standard wedge. Common spectra on successive films were used for standardizing one film with respect to another. The assumption was made that the heights of the peaks on the photometer trace, properly calibrated, were proportional to the integrated reflexions of the corresponding spectra. This procedure, although it has little theoretical justification, is found empirically to lead to no inconsistencies, and is certainly a much better approximation to the truth than can be obtained by merely visual estimation. The intensities so measured are, of course, on an arbitrary scale.

When an arrangement was found that gave the same general relative rise and fall of intensities as those observed it was possible to express the observed intensities on an approximately absolute scale for the purpose of carrying out the Fourier synthesis. It may here be pointed out that if the wrong absolute scale has been chosen the only effect on the Fourier projection will be to make the relation of the peaks to the background wrong. The projection would be useless for an electron count. On the other hand, the positions of the peaks will not be appreciably affected and these are all that are required for a structure determination. If the relation between weak and strong integrated reflexions is wrongly given, this may have some effect on the positions of the peaks in the projection, but it will not be a large effect unless the error is very considerable. A systematic error with increasing order of spectrum will have the effect of a false temperature effect and will alter the sharpness of the peaks, but not appreciably the positions of their maxima.

STRUCTURE FACTOR CALCULATIONS

The co-ordinates of a point in a general position for the space group Pbn are

$$(x, y, z), \quad (\tfrac{1}{2} - x, \tfrac{1}{2} + y, z), \quad (\tfrac{1}{2} + x, \tfrac{1}{2} - y, \tfrac{1}{2} + z), \quad (\bar{x}, \bar{y}, \tfrac{1}{2} + z).$$

The general structure factor $F(hkl)$ is given by

$$F(hkl) = \sum_j f_j(A_j + iB_j).$$

The structure factors used in this work are $F(hk0)$ and $F(0kl)$:

$$\begin{aligned} \text{For } F(hk0) \text{ when } h+k=2n, & \quad \begin{cases} A = 4 \cos 2\pi hx \cos 2\pi ky, \\ B = 0, \end{cases} \\ \text{when } h+k=2n+1, & \quad \begin{cases} A = -4 \sin 2\pi hx \sin 2\pi ky, \\ B = 0. \end{cases} \end{aligned}$$

$$\begin{aligned} \text{For } F(0kl) \text{ when } k = 2n \text{ and } l = 2n, & \quad \begin{cases} A = 4 \cos 2\pi ky \cos 2\pi lz, \\ B = 4 \cos 2\pi ky \sin 2\pi lz, \end{cases} \\ \text{when } k = 2n + 1 \text{ and } l = 2n + 1, & \quad \begin{cases} A = -4 \sin 2\pi ky \sin 2\pi lz, \\ B = 4 \sin 2\pi ky \cos 2\pi lz. \end{cases} \end{aligned}$$

f_s is the atomic scattering factor of the atom s in the unit cell, whose co-ordinates are (x_s, y_s, z_s) , and the summation is to be taken over all the atoms in the unit cell. For carbon, the atomic scattering factors given by Robertson (1935), which were derived from measurements on anthracene at low temperatures, were used. For nitrogen and oxygen the values tabulated by James & Brindley (1931) were used as basis, but they were reduced in the same ratio as that between Robertson's values for carbon and those of James & Brindley, in order to make some allowance for thermal motion.

The provisional structure was adjusted until the observed structure factors of the strongest spectra of the type $hk0$ agreed with those calculated from the above formulae using the assumed co-ordinates. In calculating the observed structure factors it was assumed that the crystals reflected as mosaics, so that the intensities were proportional to the squares of the structure factors. It was found possible to get fair agreement with two types of packing, and these were now tested by means of a Fourier projection.

THE FOURIER SYNTHESIS

The projection of the structure on the plane (001) , which uses spectra of the type $hk0$, is the most suitable for the purpose. The density $\sigma(x, y)$ of the projection at the point (x, y) on the ab plane is given by

$$\sigma(x, y) = \frac{1}{A(ab)} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} F(hk0) \exp \left[-2\pi i \left(\frac{hx}{a} + \frac{ky}{b} \right) \right],$$

where $A(ab)$ is the area of the side ab of the unit cell. The projection has a centre of symmetry, although the structure as a whole has not, and the series in the case takes the form

$$\sigma(x, y) = \frac{1}{A(ab)} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \pm |F(hk0)| \cos 2\pi \left(\frac{hx}{a} + \frac{ky}{b} \right).$$

The observed values of $|F(hk0)|$ are used as the coefficients in the series, the sign appropriate to any given coefficient being determined from the provisional structure. The two types of packing that gave possible *magnitudes* to the calculated structure factors for the strong spectra gave opposite signs to some of them and the Fourier projection was made with each set of signs in turn, the method used being substantially that described by Lipson & Beevers (1936). One set of signs gave a projection corresponding to nothing physical, but the other gave very clear outlines of the molecules. From this projection, more accurate values of the x and y parameters were determined, a recalculation of the structure factors was made, which

resulted in a change of sign of some of the smaller coefficients, and a second projection was carried out. A third and final projection was made as a result of the second, and this is illustrated in figure 1, which may be considered in conjunction with the key diagram figure 2.

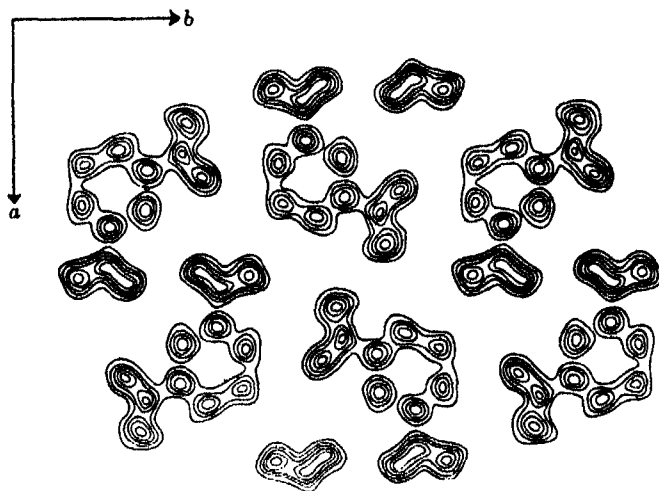


FIGURE 1. Fourier projection of the structure in the c direction on the ab plane. Contours at intervals of approximately 1 electron per sq.Å. The outer line is the 2-electron contour.

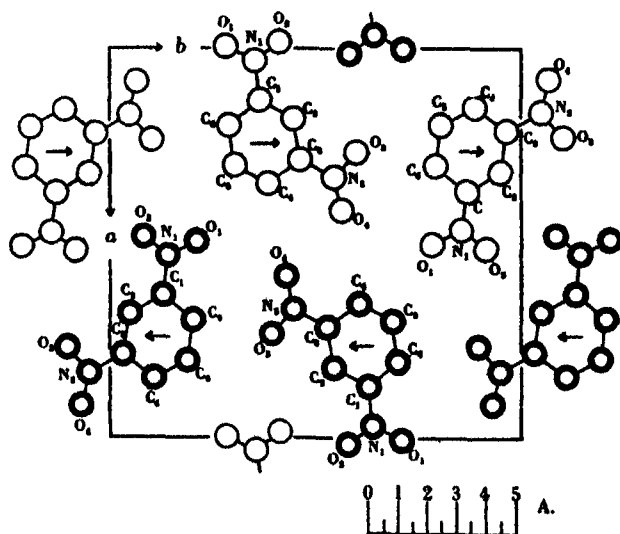


FIGURE 2. Projection of the structure in the c direction on the ab plane. The molecules with double circles are $\frac{1}{2}c$ above or below those with single circles.

The agreement between F_{observed} and $F_{\text{calculated}}$ (for the third projection) is good, with no serious discrepancies. Table 3 shows the agreement between the observed and calculated F values for the spectra $hk0$ and $0kl$.

ESTIMATION OF PARAMETERS

In the projection on the *c*-face the atoms are all well resolved, and it is possible to fix the *x* and *y* parameters with considerable accuracy. From these, in conjunction with the known dimensions of the molecule, it is possible to draw certain conclusions. The lines joining opposite carbon atoms of the benzene ring pass very nearly through one point in the middle of the ring. This suggests that the hexagon is regular, but the different lengths of the sides indicate that the plane of the ring is not parallel to the (001) plane. The lengths of the sides nearly parallel to the *a*-axis are 1.41 Å, but the other sides have in the projection a length 1.33 Å. If the ring is regular, its plane must make an angle of about 20° with the *b*-axis and be nearly parallel to the *a*-axis. This result is in agreement with the results of Banerjee & Bhattacharjya, of which it has been possible to obtain only an abstract, who from observations of the magnetic anisotropy of the crystals conclude that the benzene ring is parallel to the *a*-axis, but is inclined at an angle of 22° 20' to the *b*-axis. In the projection the C₁N₁ distance is 1.40 Å, the C₃N₂ distance is 1.33 Å. Allowing for the above-mentioned tilt of the molecule of 20° to the *b*-axis, the CN distances become 1.42 Å, if it is assumed that the CN link lies in the plane of the benzene ring. These CN distances are shorter than the CN distances observed by James *et al.* (1935) for para-dinitrobenzene and van Niekerk (1943) for 4:4'-dinitrodiphenyl, the average value of which is 1.54 Å. The discrepancy can be explained by assuming the CN link to be inclined at an angle of about 15° to the plane of the ring. In figure 2, if the directions of the arrows on the molecules indicate a tilt downwards, then N₂ is below the plane of the ring and N₁ could be either above or below C₁.

TABLE 1. INTERATOMIC DISTANCES IN THE NITRO GROUP (IN Å)

	C-N	N-O ₁	N-O ₂	O-O
James <i>et al.</i> (1935), para-dinitrobenzene	1.53	1.25	1.10	2.14
van Niekerk (1943), 4:4'-dinitrodiphenyl	1.56	1.21	1.14	2.00

The NO distances in the projection are all different, and this suggests that the nitro groups do not lie in the (001) plane. James *et al.* (1935) and van Niekerk (1943) found the nitro group to be unsymmetrical, and the values of the interatomic distances obtained by them are given in table 1. In neither case were the observations accurate enough to make this lack of symmetry entirely certain. A symmetrical nitro group has been observed in some recent work by Mr D. H. Saunder in this laboratory, not yet published, on a molecular complex of 4:4'-dinitrodiphenyl and 4-hydroxydiphenyl, in which the nitro groups are found to lie across a mirror plane of symmetry. Various possible N-O distances were assumed, and the corresponding *z*-parameters calculated for both possible positions of each oxygen atom in relation to the corresponding nitrogen atom, and for both possible positions of N₁ relative to C₁. These *z* parameters were checked by comparing the calculated and observed *F* values for

the $0kl$ spectra. The best agreement was obtained with the nitro group symmetrical, the N-O distance being 1.20 Å, as shown in figure 4*a* and *b*. It was found that the angle ONO was then 130° and the O-O distance 2.17 Å. These values are in good agreement with those obtained by James *et al.* (1935) and van Niekerk (1943).

An examination of figure 3, which shows the positions of the atoms in the crystal projected on to the bc plane, shows that the overlap of the molecules is such that no very good resolution is to be expected, particularly as the number of spectra available for the projection is small, owing to the small value of the c spacing and to space-group restrictions. Similar remarks apply to the projection on the ac plane. However, a projection of the electron density on the bc plane was made, using visual estimates of the intensities of the spectra $|F(0kl)|$. In this projection there is no centre of symmetry, so it was necessary to calculate the phases in order to carry out the summation. No definite values for the parameters could be obtained from this projection, but the large unresolved peaks were quite consistent with the arrangement of atoms shown in figure 3. It is worth noting that the molecules lie very nearly in the planes (021) and $(0\bar{2}1)$, which would account for the strong thermal ghost of the layer lattice type that accompanies the corresponding spectra.

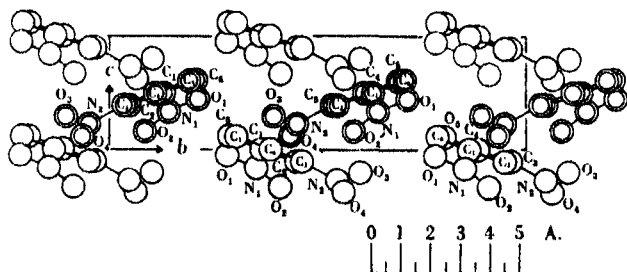


FIGURE 3. Projection of the structure in the a direction on the bc plane. The molecules with double circles are $\frac{1}{2}a$ above or below those with single circles.

The parameters of the atoms, expressed as fractions of the lattice translations, are shown in table 2.

TABLE 2. PARAMETERS OF THE ATOMS AS FRACTIONS OF THE CORRESPONDING LATTICE TRANSLATIONS

	x	y	z		x	y	z
C ₁	0.135	0.367	0	N ₁	0.032	0.354	-0.174
C ₂	0.180	0.452	-0.109	N ₂	0.333	0.546	-0.283
C ₃	0.285	0.463	-0.109	O ₁	-0.003	0.284	-0.063
C ₄	0.347	0.388	0	O ₂	-0.017	0.411	-0.345
C ₅	0.300	0.304	0.109	O ₃	0.268	0.601	-0.213
C ₆	0.197	0.292	0.109	O ₄	0.416	0.564	-0.388

DESCRIPTION OF THE MOLECULE

From the calculations already described, the molecule has the following dimensions: C-C = 1.41 Å, C-N = 1.54 Å, N-O = 1.20 Å, O-O = 2.17 Å.

The size and shape of the molecule is given in figure 4 *a* and *b*.

Note. In figure 4 *a* the angle of 6° which the projected line joining C_1N_1 makes with the line C_1C_4 may be due to N_1 and O_2 being unresolved in the projection (see figure 1) and may not be a real effect.

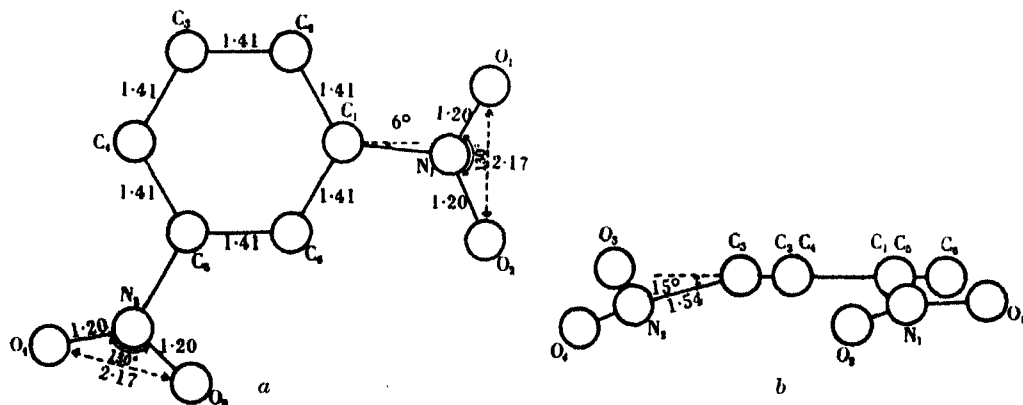


FIGURE 4. *a*. The molecule of meta-dinitrobenzene projected on to the plane of the benzene ring. *b*. The molecule of meta-dinitrobenzene projected on to a plane normal to that of the benzene ring.

It must be understood that this picture of the molecule of meta-dinitrobenzene has been derived from only one projection (that on the *ab* plane), and the size and shape of the nitro group has been largely assumed from the results obtained for other nitro compounds. This description refers to the molecule as a unit of the crystal structure, and a free molecule may not necessarily have the same configuration.

THE PACKING OF THE MOLECULES IN THE CRYSTAL

There are 4 molecules in the unit cell which are derived from a single molecule by operation of the glide planes. The diagrams are labelled so that atoms derived from one another by a lattice translation, or by the operation of a glide plane are denoted by the same letter.

The packing of the molecules in the crystal is such that each oxygen atom approaches one CH unit rather closely (the distances of approach varying from 2.95 to 3.25 Å) and the other CH units at a much greater distance (these distances varying from 3.55 Å upwards). There seems to be no tendency for an oxygen atom to pack itself at equal distances from neighbouring CH units. These shortest distances can be compared with the O-CH distance of 3.3-3.4 Å obtained by James

TABLE 3. COMPARISON OF OBSERVED AND CALCULATED F VALUES

spectrum	$4 \times F_{\text{obs.}}$	$F_{\text{calc.}}$	spectrum	$4 \times F_{\text{obs.}}$	$F_{\text{calc.}}$
200	23.6	-19.2	510	14.8	-12.8
400	34.0	46.8	520	33.6	-34.4
600	50.8	42.8	530	13.2	10.0
800	abs.	0	540	12.4	-20.4
10, 00	10.0	12.4	550	abs.	-2.0
12, 00	abs.	8.4	560	16.4	16.8
020	27.6	32.0	570	16.4	25.2
040	27.2	-32.0	580	15.2	22.4
060	33.2	-42.0	590	abs.	-4.8
080	17.0	-16.0	610	abs.	-7.2
010, 0	5.6	4.8	620	14.0	-16.0
012, 0	18.6	-22.8	630	3.9	-1.2
110	10.8	-13.6	640	abs.	5.2
120	25.6	18.8	650	abs.	5.2
130	43.6	53.2	660	abs.	-0.8
140	34.4	-45.2	670	5.2	5.2
150	10.8	-5.2	680	abs.	1.2
160	16.8	13.6	6, 12, 0	6.8	-11.6
170	6.4	-5.2	710	34.0	-36.8
180	10.4	10.8	720	16.0	23.2
190	10.0	-17.2	730	8.0	5.2
1, 10, 0	10.0	19.2	740	8.8	-8.0
1, 11, 0	10.8	5.2	750	abs.	5.2
1, 13, 0	6.4	-7.2	760	abs.	-2.4
210	26.0	-30.0	770	9.2	16.0
220	23.2	-30.4	780	abs.	-2.8
230	31.2	-37.6	810	abs.	1.6
240	12.4	-12.8	820	10.0	-8.8
250	22.0	-21.6	830	9.2	14.4
260	abs.	-6.4	840	12.8	-15.2
270	15.2	-16.4	850	abs.	4.0
280	9.6	8.0	860	14.8	-9.2
290	10.8	-9.6	870	16.4	-22.4
2, 10, 0	abs.	-4.4	880	14.8	16.8
2, 12, 0	12.8	10.8	890	16.0	-18.8
2, 13, 0	9.6	14.4	910	abs.	3.2
310	53.2	-59.6	920	abs.	5.2
320	abs.	6.8	930	5.2	7.2
330	25.2	22.0	940	4.8	6.0
340	12.8	-13.6	950	abs.	-2.4
350	14.0	13.6	960	15.6	-10.4
360	43.2	-51.6	970	abs.	8.8
370	17.6	19.2	980	10.0	13.6
380	abs.	-2.4	10, 10	abs.	-2.4
390	7.2	-13.2	10, 20	abs.	-8.8
3, 10, 0	abs.	7.2	10, 30	10.0	11.6
410	5.6	-1.2	10, 40	12.4	20.4
420	21.2	-22.0	10, 50	22.4	25.2
430	10.0	-5.6	10, 60	abs.	-4.0
440	abs.	2.0	10, 70	12.0	13.6
450	15.2	16.8	10, 80	13.2	-14.8
460	8.4	-14.8	11, 60	9.6	11.6
470	abs.	-2.8	11, 70	6.8	8.0
480	6.8	1.2	11, 80	8.4	3.2
490	8.0	-3.6	12, 60	6.8	-11.6
4, 10, 0	10.8	19.6	13, 10	7.2	10.8
			13, 20	4.8	-10.8

Table 3 (continued)

spectrum	$F_{\text{obs.}}$	$F_{\text{calc.}}$	δ
021	60	67.6	278°
041	25	17.3	35°
061	25	34.8	290°
081	10	11.7	198°
0, 10, 1	12	19.5	259°
0, 12, 1	abs.	10.2	62°
002	6	6.6	76°
022	38	18.8	177°
042	26	7.1	344°
062	8	15.7	289°
082	12	18.0	288°
0, 10, 2	15	9.9	76°
023	abs.	9.6	222°
043	8	11.6	196°
063	13	16.2	98°
083	9	14.1	98°

et al. (1935) with para-dinitrobenzene and the O-CH distances of 2.9–3.2 Å and 3.4–3.6 Å obtained by van Niekerk (1943) with 4:4'-dinitrodiphenyl.

The closest approach of CH units to one another is 3.6 Å, which is in agreement with other work done on aromatic compounds.

The closest approach of neighbouring oxygen atoms varies from 3.2 to 3.7 Å, which distance is considerably greater than the closest distance of approach for O-O in inorganic compounds, which is 2.7 Å, but agrees with the values obtained by James *et al.* and van Niekerk for aromatic nitro compounds.

The forces binding the crystal together are probably van der Waals's forces, although it seems possible that there may be some attractive force between neighbouring O and CH units, which helps to bind the crystal together.

The structure is a compact one. The crystal has a high density, 1.57 g./c.c., and, like other organic crystals, is fairly soft and easily cut. It is unlike para-dinitrobenzene and 4:4'-dinitrodiphenyl in this respect, although the main binding forces of the crystal appear to be of the same type.

I wish to express my thanks to Professor R. W. James, of the University of Cape Town, for his constant help and encouragement throughout the course of this work, and in particular for the use of unpublished data on the refractive indices and the polarity of the crystal.

REFERENCES

- Banerjee & Bhattacharjya 1938 *Sci. & Cult.* **4**, 60.
 Banerjee & Ganguly 1940 *Indian J. Phys.* **14**, 231.
 Hendricks & Hibbert 1931 *Amer. Chem. Soc.* **53**, 4280.
 Hertel 1930 *Z. Phys. Chem. B*, **7**, 188.
 James & Brindley 1931 *Phil. Mag.* **12**, 81.

- James, King & Horrocks 1935 *Proc. Roy. Soc. A*, **153**, 225.
Lipson & Beevers 1936 *Proc. Phys. Soc.* **48**, 772.
Lonsdale 1942 *Proc. Phys. Soc.* **52**, 314.
Lonsdale & Orelkin 1934 *Proc. Roy. Soc. A*, **144**, 630.
Robertson 1935 *Proc. Roy. Soc. A*, **50**, 106.
Steinmetz 1915 *Z. Kristallogr.* **54**, 467.
van Niekerk 1943 *Proc. Roy. Soc. A*, **181**, 314.
-

Measurement of the ultimate pressures of oil-diffusion pumps

BY J. BLEARS, B.Sc.(ENG.)

(Communicated by C. R. Burch, F.R.S.—Received 17 March 1945)

By using two ionization gauges to measure the pressure in the same vessel it is shown that the vapours of diffusion-pump oils are absorbed in the gauges by two processes, one transient, the other continuous. At the steady state the equilibrium in the electrode space of each gauge is dynamic, and the equilibrium pressure depends upon the dimensions of the tube connecting the gauge to the vacuum system. From the steady-state readings the ultimate pressure of the pump, the partial pressures of the reaction products and the errors of both gauges are calculated for the particular case of Apiezon 'B' oil.

The ultimate pressures obtainable by using a number of oils in fractionating and non-fractionating pumps have been investigated, and it is shown that the values obtained with different kinds of oil in the same pump do not differ widely. The lowest pressures obtainable by oils alone are much higher than is generally believed.

♦ I. INTRODUCTION

Since the introduction of the oil-diffusion pump in 1928 there have been many attempts to measure the lowest pressure it could produce. Most of this work has been carried out in industrial laboratories, and only three detailed papers (Hickman 1936*a, b*; Jaekel 1942) appear to have been published. In each of these papers, as well as in much of the unpublished work, ionization gauges were used to measure the pressure, notwithstanding frequent statements in the literature (e.g. Farkas & Melville 1939) to the effect that measurements made with this type of gauge may be entirely invalidated by reactions between the gas and the incandescent filament. There is thus an uncertainty in the interpretation of ionization gauge measurements quite apart from that occasioned by the lack of data on probability of ionization.

In his early work on this subject, Burch (1929) considered that the true ultimate pressure of a pump was the lowest value recorded after the gauge had been thoroughly outgassed, since only then could it be assumed that gas from the electrodes was not making a contribution towards the total pressure. Several years later, Hickman (1936) pointed out that the initially low value obtained immediately after outgassing was not maintained. He found that after degassing, the gauge reading passed through

a minimum and then rose to an asymptote having the same value as that to which it fell if no heat treatment whatever were applied. Since this asymptote was reproducible, Hickman was led to the conclusion that its value gave the ultimate pressure of the pump.

By using two gauges to measure the pressure in the same vessel, in the manner shown by figure 1, the author has demonstrated that it is possible for an ionization gauge to indicate consistently a pressure of the order 10 times that at the asymptote (Blears 1944), and has concluded that there is a continuous absorption process going on in the gauge, which is quite independent of the transient effect reported by Hickman.

Figure 2 (which is reproduced by courtesy of the editors of *Nature*) gives the curves obtained from the gauges shown in figure 1. It illustrates the low value used by Burch, the asymptote observed by Hickman, and the high value obtained from the gauge directly inserted in the vacuum system.

Apparently the possibility of a continuous absorption process in the ionization gauge has not previously been considered. It has generally been assumed that after the initial absorption period, the errors produced by reactions in the gauge are negligible, and the equilibrium in the system is static. The only way in which it has been possible to account for the difference between the pressure indicated by the normal gauge and by the one having very free access to the vacuum system, has been to assume that there is a continuous flow of residual gas into the gauges. On this view the equilibrium is dynamic rather than static, and accordingly the apparent pressure indicated by the gauge is a function of the size of tube connecting it to the vacuum system.

From this fundamental point it follows immediately that a true estimate of the pressure in a vessel containing reacting or absorbable gases can only be made when the pressure drop down the connecting tube is either negligible or calculable.

The purpose of this paper is first to show that it is possible to estimate the errors of an ionization gauge caused by gas absorption, and secondly to give the results obtained when a gauge designed to keep these errors very small was applied to the measurements of the ultimate pressures produced by a number of organic pump fluids in two types of oil-diffusion pump.

II. THE EQUILIBRIUM IN A VACUUM SYSTEM

(1) *Apparatus*

The apparatus used for investigating the characteristics of the ionization gauge is shown in figure 1. A 10 l. steel bell jar was evacuated by a standard pumping plant consisting of a rotary pump and two oil-diffusion pumps in cascade (Metrovac D.R. 1, 02 and 03). The diffusion pump fillings were Apiezon 'A' and 'B' oils respectively. The internal baffles of the 03 pump had been removed and replaced by a large internal baffle in the bell jar itself. Two ionization gauges having identical electrode systems were used to measure the pressure, one being mounted on a tubula-

tion in the normal manner, and the other being removed from its glass envelope and mounted directly inside the bell jar. These gauges have been called the 'normal' and the 'high-speed' gauges respectively.

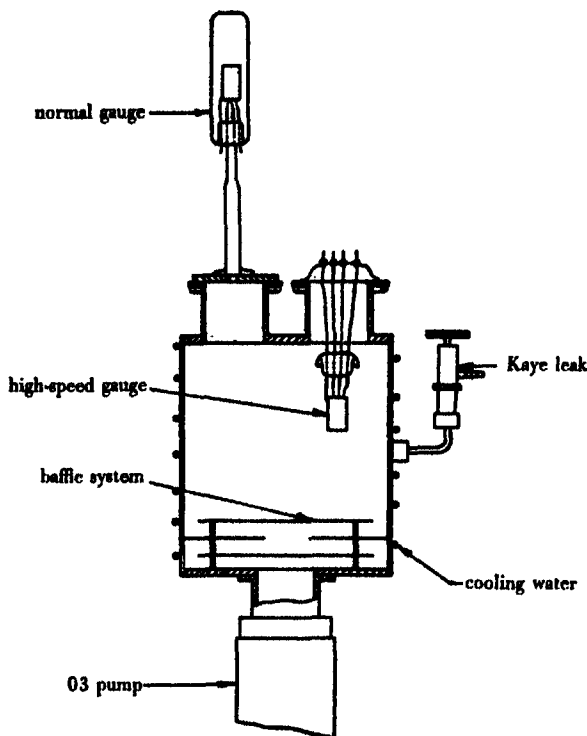


FIGURE 1. Apparatus for experiments on ionization gauges.

The gauges were controlled by automatic circuits substantially of the type devised by Ridenour & Lampson (1937), and the positive ion currents were measured by a d.c. amplifier having a sensitivity of 10,000 mm. per μA . The temperature of the bell jar was controlled by heating the cooling water, and measured by mercury thermometers reading to 0.1°C .

(2) *Experimental*

(a) *Nature of the residual gas.* It was observed that the pressure indicated by the gauges was extremely temperature sensitive, a fact which could mean either that the residual gas was a vapour, or that it was driven from the walls as an exponential function of temperature. The latter hypothesis was tested by admitting a series of gases and vapours into the bell jar through the needle valve. Water vapour, carbon dioxide and hydrogen were admitted first, since these were the most probable surface absorbates, but it was found that each of these produced equal incremental pressure changes in both gauges, whereas if one of them had been identical with the residual

gas it would have produced a curve Pg_1 versus Pg_2 with the same slope as that obtained by heating the bell jar. Later N_2 , CO , C_2H_6 , C_3H_8 were admitted with a similar effect, and it was concluded on this evidence that the substance absorbed by the normal gauge could be none of these. It could therefore hardly be anything other than the vapour of the working fluid, and that this was the case was confirmed

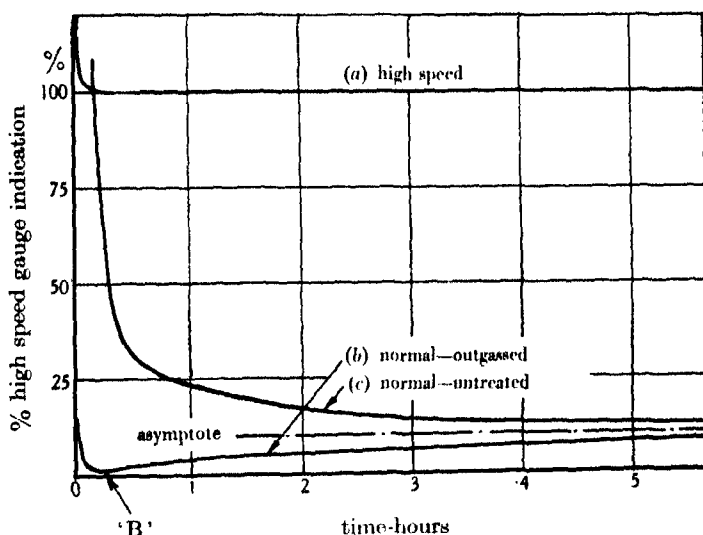


FIGURE 2. Relative indications of normal and high-speed ionization gauges.

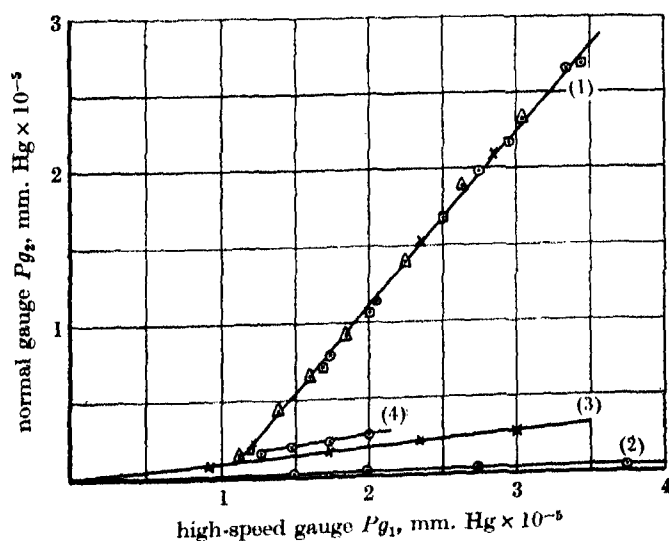


FIGURE 3. Equivalent nitrogen pressures indicated by normal and high-speed gauges after admission of gases and vapours. (1) Air, N_2 , CO_2 , C_3H_8 ; (2) residual gas immediately after baking normal gauge; (3) residual gas 24 hr. after baking; (4) Apiezon 'B' oil directly distilled in bell jar. Note. Pressure of residual gas was changed by heating bell jar.

by distilling a small quantity of Apiezon 'B' oil from a test-tube mounted directly inside the bell jar. Though this experiment is open to criticism, the slope of the curve P_{g_1} versus P_{g_2} is almost identical with that obtained after heating the bell jar, and leaves little room for doubt that the residual substance is the vapour of the pump oil. Curves for P_{g_1} versus P_{g_2} obtained under various conditions are given in figure 3.

(b) *Pressure-temperature relationships.* The normal gauge was thoroughly out-gassed by torching repeatedly until the minimum value (i.e. that at 'B', figure 2) remained constant after two or more consecutive torchings. The temperature of the bell jar was kept constant throughout this heat treatment and, when consistent readings had been obtained from the normal gauge, the 'B' value was noted together with the temperature and the corresponding reading of the high-speed gauge. A second 'B' value was obtained at the same time by momentarily switching off the high-speed gauge. This procedure was repeated at four different temperatures.

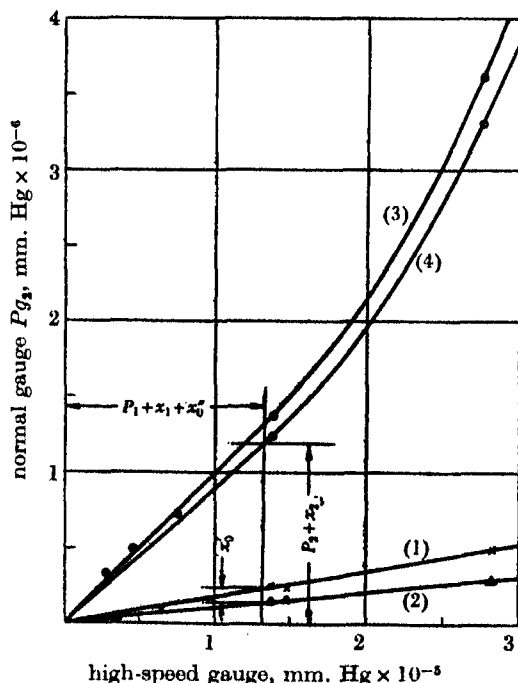


FIGURE 4. Normal gauge versus high-speed gauge indication. Pressure increased by raising bell-jar temperature. (1) 'B' points immediately after baking; (2) 'B' points immediately after baking with high-speed gauge shut off; (3) asymptotes 16-24 hr. after baking; (4) asymptotes 16-24 hr. after baking with high-speed gauge shut off.

The results are plotted in figures 4 and 5. Figure 4(1) shows the equivalent nitrogen pressure indicated by the normal gauge plotted against that given by the high-speed gauge; figure 4(2) shows the drop in the indication of the normal gauge caused by switching off the high-speed gauge momentarily. Figure 5(1) and (2)

show the same pressures plotted on log-linear co-ordinates to a base of temperature. The linearity of these curves leads to the conclusion that the indications of both gauges are proportional to the pressure in the bell jar over quite wide intervals, since, as has been shown, the residual substance is a vapour.

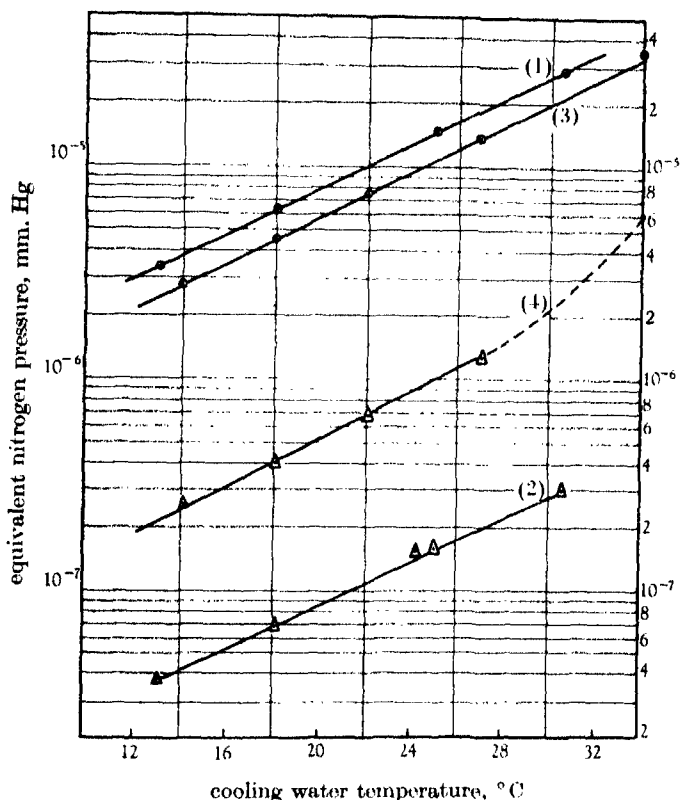


FIGURE 5. Curves showing that indications of both gauges are proportional to bell-jar pressure for any consistent method of operation. (1) High-speed gauge after 16-24 hr. running; (2) normal gauge 'B' points after 16-24 hr. running; (3) high-speed gauge after 40-48 hr. running; (4) Normal gauge asymptotes 16-24 hr. after baking.

After this experiment both gauges and pumps were allowed to run for a further 16 hr. so that the reading of the normal gauge approached fairly closely to the asymptote. A second series of readings at four different temperatures was then obtained, between 1 and 2 hr. being allowed at each temperature for conditions to reach equilibrium. These results are also plotted in figures 4 and 5. It will be observed that in this case the ratio P_{g1}/P_{g2} is not nearly so great as in the previous experiment, and further that the linearity is not maintained at pressures greater than 2 or 3×10^{-5} (on the high-speed gauge). The pressure-temperature curve for the high-speed gauge was completely linear, that of the normal gauge showed non-linearity at temperatures greater than 28°C . From this fact it was concluded that

at higher pressures the normal gauge was no longer able to absorb the same fraction of incident molecules as it could at lower pressures, and analysis was therefore restricted to the low-temperature parts of the curves where linear relationships held.

(3) *Flow of gases and vapours in the system*

Figure 6 gives a diagrammatic picture of the system and indicates the flow of vapours and reaction products at the steady state. From this diagram it is possible to put forward an explanation of the observed phenomena.

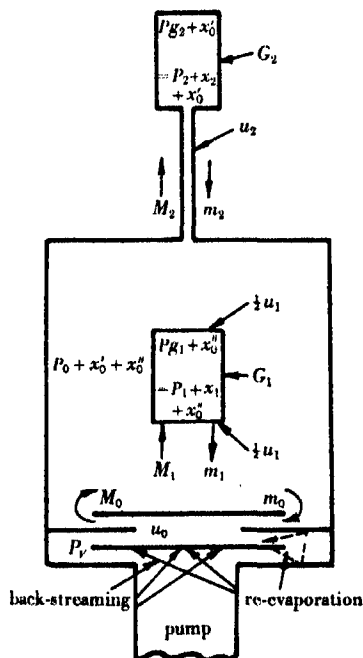


FIGURE 6. Processes and pressures in the system when both gauges are working simultaneously.

Because of vapour absorption, a mass of vapour M_0 is evaporated from the underside of the baffle in each second, and this evaporation supplies M_1 and M_2 , the masses of vapour absorbed per second by the gauges G_1 and G_2 . The masses of vapour entering G_1 and G_2 can be accounted for partly by thermal decomposition and partly by absorption. If m_1 and m_2 are the masses of reaction products formed per second, the resultant partial pressures of reaction products will be x_1 and x_2 respectively, x_1 and x_2 being determined by the rate of the reaction, the conductances of the electrodes and evacuating tube, and the speed of the pump.

(4) *Resultant pressures*

When both gauges are running, the pressure in the bell jar is made up from four components: (a) a residual permanent gas pressure, which was less than 0.25 % of the high-speed gauge indication and has therefore been neglected; (b) a vapour

pressure P_0 slightly less than P_v , the ultimate pressure of the pump (because of the pressure drop suffered by the vapour in flowing through the baffles); (c) x'_0 and x''_0 the resultant pressures in the bell jar caused by x_1 and x_2 in the gauges. The values of x'_0 and x''_0 will of course be less than x_1 and x_2 by calculable amounts.

Under the same circumstances the pressure in the electrode space of G_1 is made up from: (a) P_1 the partial pressure of unconsumed oil vapour, (b) x_1 the partial pressure of reaction products formed by thermal decomposition in G_1 , (c) x''_0 the equilibrium pressure in the bell jar of reaction products formed in G_2 .

Similarly, the pressure in G_2 is made up from P_2 of unconsumed oil vapour, x_2 of reaction products formed in its own envelope, and x'_0 of reaction products formed in G_1 .

It has been shown that both gauges respond equally to gases such as methane and propane, therefore if it is assumed that the reaction products take the form of these or similar gases, it follows that both gauges will respond equally to reaction products. Thus, when the high-speed gauge is momentarily shut down the partial pressure x'_0 will disappear from the system, and the normal gauge G_2 will experience a change of pressure of this value. (At the same time the vapour pressure in the bell jar P_0 will increase almost to P_v since M_1 is no longer being absorbed by the high-speed gauge, but it may be shown that this effect is negligible at the 'B' values.)

Thus, the partial pressure of reaction products formed by the high-speed gauge can be deduced by observing the change of pressure registered by the normal gauge when the high-speed gauge is switched off.

The value of x_2 , the partial pressure of reaction products formed in G_2 , cannot be obtained so readily, since throwing off this gauge momentarily does not produce a measurable change in the indication of the high-speed gauge. It therefore becomes necessary to make the following assumptions:

(a) That at the asymptotes when all adsorption on the glasswork has ceased, both gauges absorb oil vapour by the same mechanism.

(b) That so long as the $\log P_0$ - T curves are linear, the mass of vapour absorbed per second, and the mass of reaction products formed per second, is proportional to the partial pressure of oil vapour in the respective electrode systems.

(5) Calculation of partial pressures of vapour and reaction products for Apiezon 'B'

A first approximation to the value of x_1 , the partial pressure of reaction products formed in the high-speed gauge, is given directly by the difference in the curves (1) and (2) of figure 4. The value measured experimentally is x'_0 , and this of course is less than x_1 by an amount which can be calculated in terms of u_1 , the conductance of the gauge electrodes, and S_p , the speed of the pump.

Figure 4 gives the value of x'_0 directly in terms of the total indication of the high-speed gauge as

$$x'_0/(x_1 + x''_0 + P_1) = 1/140. \quad (1)$$

Calculation yields the value of x_1 :

$$x_1/(x_1 + P_1) = 1/96. \quad (2)$$

This result means that the ionization of reaction products accounts for only just over 1 % of the total indication of the high-speed gauge compared with the 0.7 % given by the direct experiment.

No simple approximation to the partial pressure of reaction products formed in the normal gauge has yet been found, but analysis based on the assumptions (a) and (b) above leads to the expressions:

$$\frac{x_2}{P_1} = \frac{1 + \frac{x_1}{P_1}}{i_1 \left[1 + \frac{U_2}{U_1} \frac{P_1}{x_1 - x_0'} \frac{S_p}{S_p + U_2} \right] - \frac{U_2}{S_p + U_2}}, \quad (3)$$

$$\frac{P_2}{P_1} = \frac{U_2}{U_1} \frac{x_2}{x_1 - x_0'} \frac{S_p}{S_p + U_2}, \quad (4)$$

where i_1, i_2 are the observed positive-ion currents, U_1, U_2 the conductances of the gauges and S_p is the speed of the pump.

By substituting numerical values in these equations it can be shown that reaction products formed in the normal gauge account for just under 20 % of its indication. Thus in neither case does ionization of reaction products account for the major constituent of the ionization current.

The partial pressure of oil vapour in the bell jar, P_0 , can be obtained by substituting equation (4) in the readily deducible expression

$$\frac{P_1}{P_0} = \frac{U_1 - U_2 P_1 / P_2}{U_1 - U_2}, \quad (5)$$

and the ultimate pressure of the pump P_v then follows from

$$\frac{P_v}{P_0} = 1 + \frac{U_1}{U_0} \left(1 - \frac{P_1}{P_0} \right), \quad (6)$$

where U_1 is the conductance of the baffle.

Substituting numerical values in these equations and expressing the results in terms of the indication of the high-speed gauge gives

$$P_{g_1} = 0.83 P_0, \quad P_v = 1.26 P_{g_1}. \quad (7)$$

Thus the high-speed gauge indication corresponds to 83 % of the bell-jar pressure and 79 % of the ultimate of the pump.

These results contrast very favourably with the indications of the normal gauge which, even at the asymptote, records only 7.2 % of the pressure in the bell jar, the remaining 92.8 % being accounted for by absorption.

The whole position as to pressures at the steady state in the various parts of the system has been summarized in table 1. P_0 has been given the arbitrary value 100 and all the other pressures are relative to this.

TABLE 1. RELATIVE PRESSURES IN DIFFERENT PARTS OF THE SYSTEM (APIEZON 'B' OIL)

ultimate pressure of the pump, P_v	100.00
bell-jar pressure, P_0	96.0
high-speed gauge indication due to oil vapour, P_1	78.7
high-speed gauge indication due to reaction products, x_1	0.83
total indication of high-speed gauge	79.53
normal gauge indication due to oil vapour, P_2	5.82
normal gauge indication due to reaction products, x_2	1.40
total indication of normal gauge	7.22
normal gauge indication immediately after baking	0.82

On a basis of these results it may be concluded that by using a gauge having free access to oil molecules, and by providing a large evaporating surface, so that the mass of vapour absorbed per second is small compared with the rate of evaporation, it is possible to use the ionization gauge to measure the pressure corresponding to the saturation vapour pressure of the oil. The errors of such a gauge are small and can be calculated, but for normal practice it is not essential to use two gauges, provided the one used satisfies approximately the conditions laid down in the following definition:

The high-speed ionization gauge is defined ideally as a gauge in which every molecule absorbed is immediately replaced, and where every molecule produced by reactions is immediately removed, so that the pressure in the interelectrode space is always equal to the pressure in the rest of the vacuum system.

It is to be noted that such a gauge would give the right answer even if no gas or vapour absorption were taking place.

Having reached this conclusion it seemed worth while to measure with a high-speed ionization gauge the ultimate pressures of oils previously tested by the normal technique. This work will now be described.

III. MEASUREMENTS ON SOME TYPICAL OILS

(1) Apparatus

A similar pumping plant to that used for investigating the characteristics of the high-speed ionization gauge was used for the series of tests now to be described. In this case, however, the cold baffles of the 03 pump were left in position and the external bell jar and baffles were dispensed with. Two types of jet system were used in the 03 pump, first a standard single-jet system, and second, the jets of a simple type of three-stage fractionating pump, specially built to fit into the 03 body.

The electrodes of the high-speed gauge used in these experiments were identical with those of the internal gauge described in § II and were mounted in a glass dome 3 in. diameter, $3\frac{1}{2}$ in. deep which was lapped to fit on the pump flange and sealed there with Apiezon 'L' grease.

The pump cooling water was fed from a constant-head apparatus and was electrically heated by an immersion heater supplied from a constant-voltage transformer. By this simple method it was found possible to maintain the temperature of the cooling water constant within 0.1°C for long periods, the incoming water supply fortunately being at practically constant temperature for periods of a few hours. This experimental arrangement enables the overall performance of the oil, pump and baffle to be assessed.

(2) Oils investigated

The Apiezon oils investigated were of recent manufacture. The Octoil 'S' had been carefully stored in a cool dark place, but had been there since 1938. The other samples were of more recent origin, the Litton oil about 1 year old, the Arochlor about 4 months, and the East Anglia 011 about the same age when the tests were carried out on them.

(3) Experimental

(a) *Pressure-time curves.* Twenty minutes after switching on the diffusion pumps readings of the positive-ion current were commenced and continued until the apparent pressure became reasonably constant. In general, a period of 20–80 hr. elapsed before this condition was obtained. Figure 7 shows the fall in pressure during the first 2 hr. of operation of Apiezon 'B', Litton oil, and Octoil 'S' in a standard 03 pump. The cooling water in each case was at a temperature of 16.0°C . Perhaps the most significant feature of these curves is their similarity; after 2 hr. all the ultimate pressures lie between limits 0.5 and 2.0×10^{-5} mm. mercury. This pressure is entirely due to oil vapour, there being no appreciable permanent gas pressure at any time after 30 min. from switching on the pumps.

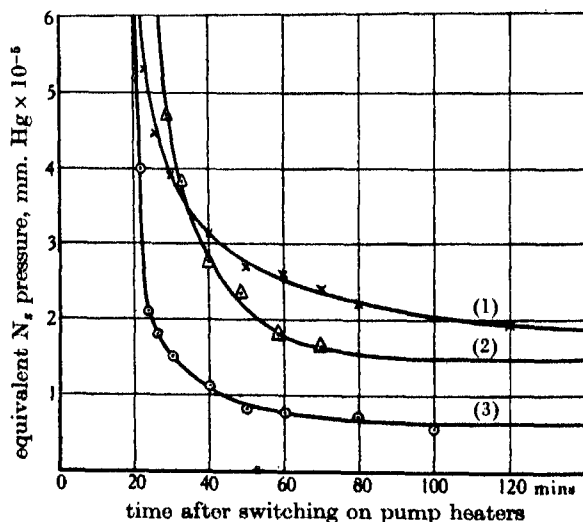


FIGURE 7. Initial values of ultimate pressure produced by a non-fractionating pump.
(1) With Apiezon 'B' oil; (2) with Litton oil; (3) with Octoil 'S'.

Readings on these oils were continued for periods up to 3 days, and the long-period results appear in figure 8. Here it will be seen that though the ultimate pressure given by 'B' oil continued to fall for the whole period, that of the other two oils showed a well-defined minimum, after 5 hr. in the case of Octoil 'S' and after 15 hr. in the case of Litton oil. After these periods the ultimate pressures of Octoil 'S' and Litton oil continued to rise, crossing the curve for Apiezon 'B' after 43 and 33 hr. respectively.

When, however, these oils were tested in the fractionating pump an entirely different state of affairs was obtained. All the oils continued to improve for the whole period of the tests, the improvement being most striking in the case of Octoil 'S' where, after 70 hr., the pressure in the fractionating pump was less than 5 % of that in the normal 03 pump after the same time. Pressure-time curves for Octoil 'S' and Apiezon 'B' in fractionating and non-fractionating pumps are given in figure 8.

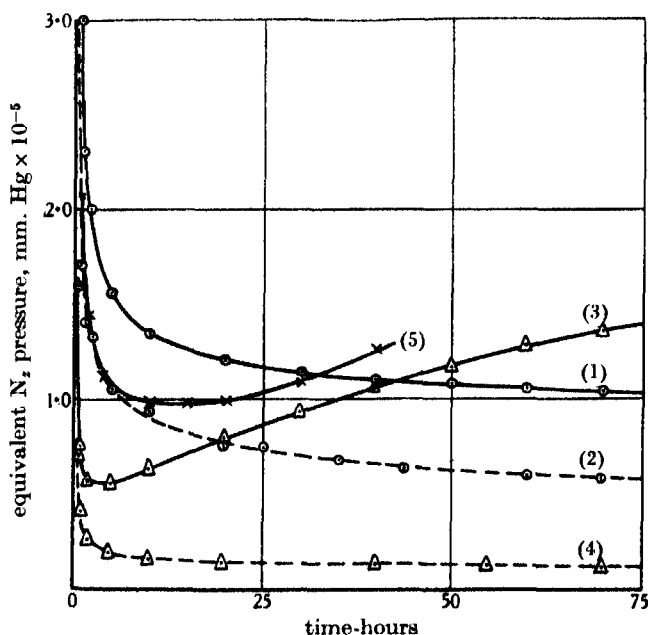


FIGURE 8. Pressure-time curves showing improvement in ultimate pressure due to fractionation. (1) Apiezon 'B' oil, non-fractionation; (2) Apiezon 'B' oil, fractionation; (3) Octoil 'S', non-fractionation; (4) Octoil 'S' oil, fractionation; (5) Litton oil, non-fractionation.

(b) *Pressure-temperature curves.* Because of the variation of pressure with time observed in the previous experiment, and of the fact that it was found necessary to wait for at least 1 hr. after changing the cooling water temperature for conditions to stabilize, it was not possible to obtain exact relationships between pressure and temperature early in the run. The procedure adopted was to run the pump for 16 hr. at constant temperature before taking the first reading, and then to vary the temperature in four distinct steps, plotting a pressure-time curve at each temperature,

finally returning to the initial temperature. If the final pressure was more than 10 % different from the initial, it was concluded that the composition of the oil had changed more than could be tolerated, and a new series of readings was made. From the pressure-time curve obtained at each temperature the pressure corresponding to that temperature was deduced and a pressure-temperature curve was then plotted on the log-linear co-ordinates.

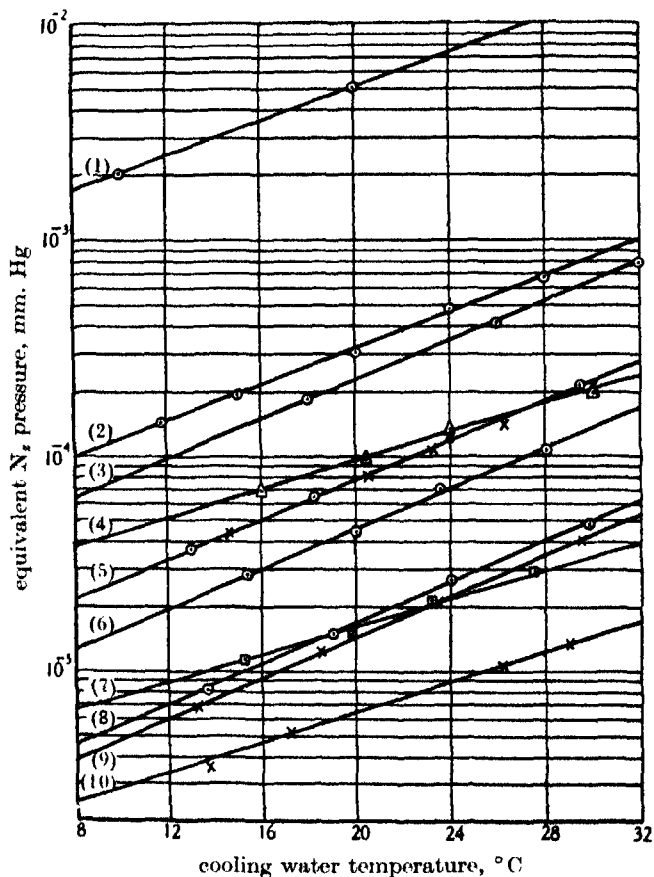


FIGURE 9. Ultimate pressures of typical fluids (non-fractionating oil pump). (1) mercury; (2) Arochlor 1254; (3) dibutyl phthalate; (4) box pump oil; (5) East Anglia 011; (6) Apiezon 'A'; (7) tricresyl phosphate; (8) Apiezon 'B'; (9) Litton oil; (10) Octoil 'S'.

In figure 9 are given the pressure-temperature relationships for nine materials which have been used as operating media for oil-diffusion pumps. For comparison purposes, a curve for mercury has been added (the 20° C value for this material was measured with the ionization gauge, but the slope is taken from the standard tables).

There is surprisingly little difference in the ultimate pressures attainable with the various oils. Omitting Arochlor and dibutyl phthalate, which are not intended for

the production of really high vacua, there is only a factor of 15 between the best and the worst oils. Ultimate pressures at 20° C for all the oils used are given in table 2.

TABLE 2. ULTIMATE PRESSURES PRODUCED BY ORGANIC DIFFUSION-PUMP OILS*

oil	pressure in mm. Hg $\times 10^{-6}$		
	measured values		values quoted in the literature
	standard 03	fractionating 03	
Octoil 'S'	6.4	2.9	0.01-1.0
Litton oil	14.0	6.6	1.0
tri-cresyl phosphate	17.0	—	—
Apiezon 'B' oil	17.0	9.2	0.1-5.0
Apiezon 'A' oil	45.0	19.0	10.0
East Anglia 011	74.0	41	—
box-pump oil	94.0	—	—
dibutyl-phthalate	225	—	100
Arochlor 1254	310	260	100

* All values obtained after running for 16-24 hr. in the pump and expressed in terms of equivalent nitrogen pressure at 20° C.

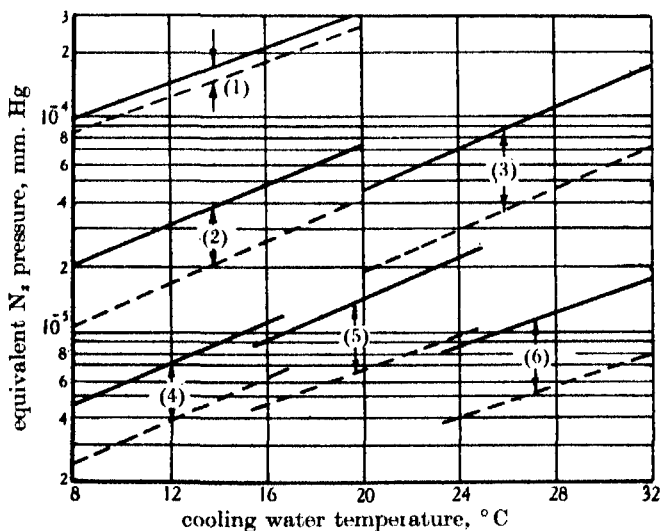


FIGURE 10. Comparative ultimate pressures in fractionating and non-fractionating pumps. (For clarity, only part of each curve is drawn.) (1) Arochlor 1254; (2) East Anglia 011; (3) Apiezon 'A'; (4) Apiezon 'B'; (5) Litton oil; (6) Octoil 'S'.

Figure 10 illustrates the improvement which was obtainable by running a number of the oils in a simple type of fractionating pump. Each of these curves was obtained 16-24 hr. after starting up the pumps and the pairs of curves, fractionating and non-fractionating, show that even with a very simple form of pump, fractionation enables

the pressure obtainable by each fluid to be reduced to a half or a third of its value in a standard type of pump. A significant point brought out by these curves, is that the single boiling-point substance Octoil 'S' fractionates to about the same extent as the supposedly wide-range hydrocarbon oils.

IV. CONCLUSIONS

When the ultimate pressures measured with a high-speed gauge are compared with those obtained with normal gauges under similar circumstances, it is found that the new values are in general 3-100 times higher than those obtained earlier during the past 15 years. On the basis of the evidence given in this paper it is concluded that the new values are a closer approximation to the truth, and that the ultimate pressures of oil-diffusion pumps are higher than was thought.

Since the production of a low pressure is dependent on the design of the pump as well as on the kind of oil used in it, it seems that further research to improve both will be needed. It should not be forgotten that there are other features of importance in the choice of an oil, for example, its thermal and chemical stability, and the rate at which it can be boiled without decomposition. All these features have to be considered before the desirability of using a new substance can be established, and it is considered that in spite of the manifest disadvantage that it responds differently to different gases, the high-speed type of ionization gauge is capable of yielding much more information of interest and value on these aspects of the problem.

It is hoped that in the near future there may be an opportunity to carry out further research on the mechanism of the continuous absorption process and on the application of the high-speed gauge to work where its short period and freedom from the effects of absorption and degassing can be used with advantage.

The author wishes to express his thanks to Mr J. Bennett for assistance with the experimental work, to Dr R. Witty and Mr G. Burrows who have discussed various aspects of the problem, to Dr C. R. Burch, F.R.S., at whose suggestion the paper was written, and finally to Sir Arthur Fleming, C.B.E., D.Eng., Director of Research and Education, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish the results.

REFERENCES

- Blears, J. 1944 *Nature*, **20**, 154.
Burch, C. R. 1929 *Proc. Roy. Soc. A*, **123**, 271.
Farkas, A. & Melville, H. W. 1939 *Experimental methods in gas reactions*. London: Macmillan.
Hickman, K. C. D. 1936a *J. Franklin Inst.* **221**, 215.
Hickman, K. C. D. 1936b *J. Franklin Inst.* **221**, 383.
Jaeckel, R. 1942 *Z. tech. Phys.* **7**, 177.
Ridenour, L. N. & Lampson, C. W. 1937 *Rev. Sci. Instrum.* **8**, 162.

The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis

By A. D. BOOTH

(Communicated by G. N. Watson, F.R.S.—Received 20 August 1945)

A brief introduction to the subject of X-ray structure analysis is followed by a discussion of various conjectures regarding the accuracy of derived atomic co-ordinates and the importance of the latter in the derivation of molecular theory.

Representing the synthesis in the form

$$D(x, y, z) = \frac{1}{V} \sum_{hkl} |F(h, k, l)| \cos \left[2\pi \left(h \frac{x}{a} + k \frac{y}{b} + l \frac{z}{c} \right) - \alpha(h, k, l) \right], \quad (1)$$

and neglecting errors of computation, two sources of inaccuracy are shown to occur: (a) experimental errors in the $|F|$ values; (b) errors due to (H, K, L) being finite.

Information as to the magnitude of the errors in the $|F|$ values has recently become available, and a comparison of two sets of experimentally determined $|F|$ values shows that the errors are independent of the magnitude of the structure factors and have a most probable value

$$\Delta e = \pm 0.6. \quad (2)$$

Examination of the shape of the atomic peaks derived from a number of Fourier syntheses shows that the radial density distribution can be closely represented by the function

$$d(r) = Ae^{-pr^2}, \quad (3)$$

where A depends on the atomic number of the particular atom and p appears to be fairly constant over a number of atoms from carbon to sulphur, a mean value being

$$p = 4.69. \quad (4)$$

A combined analytical-statistical analysis leads to the relation

$$\epsilon < 90.8 \Delta e / N \sqrt{V(\lambda p)^{\frac{1}{2}}}, \quad (5)$$

where ϵ is the most probable error in the co-ordinate, N is the atomic number of the particular atom, V is the volume of the unit cell in \AA^3 , and λ is the wave-length corresponding to the smallest spacings observed. Taking the values of Δe and p given in (2) and (4) and considering a carbon atom in a unit cell of volume

$$V = 583 \text{\AA}^3, \quad (6)$$

equation (5) leads, when all the information obtainable with copper K_α radiation is used, to the value,

$$\epsilon < 0.0027 \text{\AA}.$$

A formula is also given for the case in which errors are proportional to the order of their parent reflexions.

The problem of finite limits of summation is dealt with in Part 2. For a simple system containing only two carbon atoms the errors, calculated as upper limits, are:

ρ	$\epsilon (\text{\AA})$
1.5	0.019
1.8	0.009
2.0	0.005

where ρ is the radius of the sphere containing the reciprocal points of all planes included in the summation.

The polyatomic case cannot be given general expression since the atomic positions form a determined system and are not subject to statistical laws. In any given structure the errors

are shown to be calculable by the following procedure. Having calculated the structure factors from the final atomic co-ordinates, a synthesis is computed using these *calculated* values as coefficients. Any terms not included in the original synthesis with experimental coefficients are similarly omitted in this new synthesis. The co-ordinates derived will, in general, deviate slightly from those used in calculating the F values, these deviations give the errors, with reversed signs, of the original co-ordinates. A trial on an actual structure shows them to have a value of about

$$\epsilon = \pm 0.02 \text{ \AA}. \quad (7)$$

It is suggested that by applying these corrections to derived co-ordinates more accurate values of the latter may be obtained, having errors approximately those of experiment given by equation (5).

Finally, the value of ' p ' is related to the quantity ' B ' defined by Debye & Waller by means of the equation

$$\frac{1}{p} = a(B + c), \quad (8)$$

where a and c are constants. This relation enables the effect of thermal agitation to be examined.

1. INTRODUCTION

There has been considerable discussion in the past two or three years regarding the accuracy with which it is possible to obtain atomic co-ordinates from the Fourier series of F values devised by Bragg (1915). The classical idea that this accuracy is the same as the resolving power for the particular wave-length used in obtaining the X-ray photographs is not upheld by many workers, and the consistency shown when like bond lengths are determined in different compounds gives some weight to their claims. For example, using copper K_α radiation having a wave-length of 1.539 Å, the theoretical resolving power, given by

$$\delta = 0.6\lambda/2 \sin \theta,$$

is 0.46 Å, whereas accuracies claimed in various structure determinations range from ± 0.01 to ± 0.05 Å (Robertson 1945). Preston (1944), while quoting the classical result, considers it probable that the results of structure determinations are more accurate than its strict application would imply.

In view of the importance which small differences in bond lengths have attained for distinguishing between various resonance interpretations of molecular structure, it was very desirable that this question of accuracy be investigated and if possible put on a quantitative basis.

This paper contains an account of an investigation into the matter leading to formulae and methods from which the accuracy of any structure determination can be estimated.

One point requires mention; all results here presented apply to structure determination by means of three-dimensional Fourier syntheses. The chief objection to these has been in the past the great labour involved in their computation, but since improved computational methods (Booth 1945) have become available it appears probable (Robertson 1945) that in future all determinations will proceed on a three-dimensional rather than a projection basis.

2. ABBREVIATIONS

The following abbreviations will be used throughout this paper:

$$\theta = 2\pi \left(h \frac{x}{a} + k \frac{y}{b} + l \frac{z}{c} \right). \quad (2.1)$$

Or if a particular point (x_r, y_r, z_r) is in question:

$$\theta_r = 2\pi \left(h \frac{x_r}{a} + k \frac{y_r}{b} + l \frac{z_r}{c} \right), \quad (2.2)$$

where a , b and c are the unit cell axes in Å,

$$\alpha = \alpha(h, k, l), \quad (2.3)$$

the phase angle for the plane (h, k, l) .

$$F = F(h, k, l), \quad (2.4)$$

the structure factor.

3. DISCUSSION OF THE PROBLEM

The electron density at any point (x, y, z) in the unit cell of a crystal is given (Lonsdale 1936) by

$$D(x, y, z) = \frac{1}{V} \sum_{h,k,l} \sum_{l=-\infty}^{\infty} |F| \cos(\theta - \alpha). \quad (3.1)$$

In the calculation of D from this expression there are three sources of error:

- (1) Errors in the experimental $|F|$ values.
- (2) Termination of the series at a finite (h, k, l) value.
- (3) Computational errors.

The third of these sources will not be dealt with here, since by means of adequate precautions it can be entirely avoided. If, as is almost inevitable in three-dimensional work, the values of $|F|$ are first obtained on a relative scale by 'eye-estimation' followed by a determination of the absolute values of a few selected reflexions, the errors of experiment will fall into two classes of which it will be shown that those of 'scale' do not affect the atomic co-ordinates. The actual number of (h, k, l) values used also depends on two factors, first, the wave-length of the X-rays used, and secondly, the state of thermal motion of the atoms in the crystal.

Before either of the two sources of error is considered it is necessary to discuss the shape of electron density distributions resulting from Fourier synthesis.

4. EXPERIMENTAL DENSITY DISTRIBUTIONS

Although electron density distributions for a large number of atoms have been calculated on the basis of modern wave-mechanical theory (Hartree 1928), these are, unfortunately, only to be had in tabular and not in explicit algebraic form, and

in any case, the distributions resulting from Fourier synthesis do not greatly resemble the plots of these ideal fields. This is most probably due to the existence of thermal agitation in any real atom, a point which will be discussed in Part II. Cox & Costain (1941) have suggested that the shape of distributions obtained by synthesis are closely represented by expressions of the form

$$d(r) = Ae^{-pr^2}. \quad (4.1)$$

To verify this the peaks for carbon, nitrogen, oxygen (Booth & Llewellyn), chlorine (Jeffrey 1942) and sulphur (Cox & Jeffrey 1942) atoms, reduced to a common A value of unity have been plotted in figure 1, together with the result of making

$$A = 1, \quad p = 4.689$$

in (4.1), showing the wide applicability of the equation even when a common ' p ' value is used for several atoms.

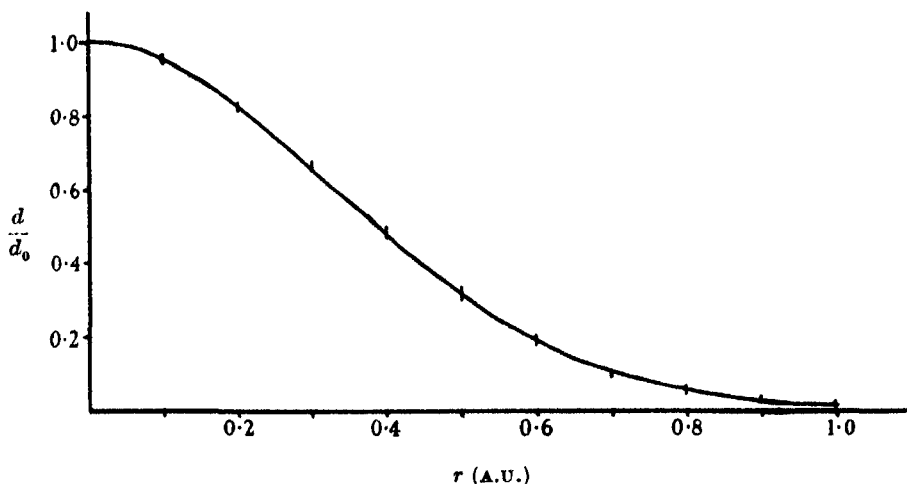


FIGURE 1. Full curve represents $\frac{d}{d_0} = e^{-4.689r^2}$. Vertical lines indicate range of experimental points for carbon, nitrogen, oxygen, chlorine and sulphur.

In order to take account of scale (4.1) must be normalized to give the right number of electrons in the atomic field. For an atom of atomic number N this leads to the relation

$$d(r) = N\left(\frac{p}{\pi}\right) e^{-pr^2}. \quad (4.2)$$

Equation (4.2) can be generalized to give the density at a point (x, y, z) near to an atom whose centre is at (x_1, y_1, z_1) :

$$d(x, y, z) = N\left(\frac{p}{\pi}\right) e^{-p[(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2]}. \quad (4.3)$$

It is convenient to record here values of the expressions

$$\left(\frac{\partial^2 d}{\partial x^2}\right), \text{ etc.}, \quad \left(\frac{\partial^2 d}{\partial x \partial y}\right), \text{ etc.},$$

at the point (x_1, y_1, z_1) .

Partial differentiation of (4.3) gives at once

$$\left(\frac{\partial^2 d}{\partial x^2}\right)_1 = \left(\frac{\partial^2 d}{\partial y^2}\right)_1 = \left(\frac{\partial^2 d}{\partial z^2}\right)_1 = -2pN\left(\frac{p}{\pi}\right)^{\frac{1}{2}}, \quad (4.4)$$

$$\left(\frac{\partial^2 d}{\partial x \partial y}\right)_1 = \text{etc.} = 0. \quad (4.5)$$

PART I

5. THE EFFECT OF EXPERIMENTAL ERRORS

Using the incomplete series, and assuming all the $|F|$ values to be perfectly accurate, equation (3.1) gives for the electron density at any point

$$D'(x, y, z) = \frac{1}{V} \sum_{-H, K, L}^{+H, K, L} |F| \cos(\theta - \alpha). \quad (5.1)$$

Now at the centre of an atomic peak the value of D' is a maximum and the co-ordinates of this point are such as to make

$$\frac{\partial D'}{\partial x} = \frac{\partial D'}{\partial y} = \frac{\partial D'}{\partial z} = 0. \quad (5.2)$$

This gives three equations of the type

$$-\frac{1}{V} \sum_{-H, K, L}^{+H, K, L} 2\pi \frac{h}{a} |F| \sin(\theta_1 - \alpha) = 0, \quad (5.3)$$

which, if they could be solved, would give the co-ordinates (x_1, y_1, z_1) of the atomic centre and show, since they do not involve $F(000)$, the only $|F|$ not determined in the relative $|F|$ estimations, that errors of scale do not effect the position of the maximum.

Next let it be assumed that the $|F|$ values are in error by small quantities ΔF and that the effect of these changes is to produce shifts $\epsilon_x, \epsilon_y, \epsilon_z$ in the co-ordinates of the resulting peak. The equations for the maximum now become

$$-\frac{1}{V} \sum_{-H, K, L}^{+H, K, L} 2\pi \frac{h}{a} (|F| + \Delta F) \sin(\theta_1 + \phi - \alpha) = 0, \quad (5.4)$$

where

$$\phi = 2\pi \left(h \frac{\epsilon_x}{a} + k \frac{\epsilon_y}{b} + l \frac{\epsilon_z}{c} \right). \quad (5.5)$$

Assuming all the errors to be small enough to neglect second order quantities (5.4) may be written

$$-\frac{1}{V} \sum_{-H, K, L}^{+H, K, L} 2\pi \frac{h}{a} |F| \sin(\theta_1 - \alpha) + 2\pi \frac{h}{a} \Delta F \sin(\theta_1 - \alpha) + 2\pi \frac{h}{a} |F| \phi \cos(\theta_1 - \alpha) = 0, \quad (5.6)$$

which, by virtue of (5.3), becomes

$$-\frac{1}{V} \sum_{-H, K, L}^{+H, K, L} 2\pi \frac{h}{a} \Delta F \sin(\theta_1 - \alpha) + 2\pi \frac{h}{a} |F| \phi \cos(\theta_1 - \alpha) = 0. \quad (5.7)$$

Using (5.5) the second term of (5.7) may be split up into three series:

$$-\frac{\epsilon_x}{V} \sum_{-H, K, L}^{+H, K, L} 4\pi^2 \frac{h^2}{a^2} |F| \cos(\theta_1 - \alpha) - \frac{\epsilon_y}{V} \sum_{-H, K, L}^{+H, K, L} 4\pi^2 \frac{hk}{ab} |F| \cos(\theta_1 - \alpha) - \frac{\epsilon_z}{V} \sum_{-H, K, L}^{+H, K, L} 4\pi^2 \frac{hl}{ac} |F| \cos(\theta_1 - \alpha), \quad (5.8)$$

which, in turn, may be written

$$\epsilon_x \left(\frac{\partial^2 D'}{\partial x^2} \right)_1 + \epsilon_y \left(\frac{\partial^2 D'}{\partial x \partial y} \right)_1 + \epsilon_z \left(\frac{\partial^2 D'}{\partial x \partial z} \right)_1, \quad (5.9)$$

and since in the region of an atomic peak

$$D'(x, y, z) = d(x, y, z),$$

equations (4.4) and (4.5) show that (5.9) reduces to

$$-2pN \left(\frac{p}{\pi} \right)^{\frac{1}{2}} \epsilon_x, \quad (5.10)$$

$$\text{whence from (5.7)} \quad \epsilon_x = -\frac{1}{NVa} \left(\frac{p}{\pi} \right)^{\frac{1}{2}} \sum_{-H, K, L}^{+H, K, L} h \Delta F \sin(\theta_1 - \alpha), \quad (5.11)$$

with similar expressions for ϵ_y and ϵ_z .

It is now necessary to consider in greater detail the series

$$\sum_{-H, K, L}^{+H, K, L} h \Delta F \sin(\theta_1 - \alpha). \quad (5.12)$$

Since all the $|F|$ values are observed in the same manner, by eye-estimation, it may be assumed that they all have the same most probable experimental error, say Δe . This being the case, the most probable value of equation (5.12) is given (Worthing & Jeffner 1943) by

$$\left\{ \sum_{-H, K, L}^{+H, K, L} h^2 \sin^2(\theta_1 - \alpha) \right\}^{\frac{1}{2}} \Delta e, \quad (5.13)$$

$$\text{which is certainly less than} \quad \left\{ \sum_{-H, K, L}^{+H, K, L} h^2 \right\}^{\frac{1}{2}} \Delta e. \quad (5.14)$$

In order to evaluate the above summation it is necessary to use the reciprocal lattice concepts introduced by Ewald (1921) and elaborated by Bernal (1926). The planes which will reflect X-rays of a given wave-length λ are those whose reciprocal points are contained in a sphere of radius 2.0 centred on the origin, the reciprocal 'unit cell' having axes

$$a^* = \lambda/a, \quad b^* = \lambda/b, \quad c^* = \lambda/c,$$

where a , b and c are the sides of the unit cell (Bernal 1926).

The number of lattice points for which h has some fixed value h_1 is given approximately (for orthogonal axes) by

$$n_1 \doteq \pi \{2^2 - (h_1 a^*)^2\} / b^* c^*, \quad (5.15)$$

$$\begin{aligned} \text{whence } \sum_{-H, K, L}^{+H, K, L} h^2 &\doteq \sum_{-H}^{+H} \pi h_r^2 \{2^2 - (h_r a^*)^2\} / b^* c^* \\ &= \frac{2\pi}{b^* c^*} \sum_1^H \{4h_r^2 - a^{*2} h_r^4\} \\ &= \frac{2\pi}{b^* c^*} H(H+1) \left\{ \frac{2}{3}(2H+1) - \frac{1}{30}(6H^3 + 9H^2 + H - 1) a^{*2} \right\}, \end{aligned} \quad (5.16)$$

whence from (5.11) and (5.16) the most probable errors are given by

$$\epsilon_x = \pm \frac{\Delta e}{NVa} \left(\frac{\pi}{p} \right)^{\frac{1}{2}} \left\{ \frac{2\pi}{b^* c^*} H(H+1) \left[\frac{2}{3}(2H+1) - \frac{a^{*2}}{30} (6H^3 + 9H^2 + H - 1) \right] \right\}^{\frac{1}{2}}, \quad (5.17)$$

with similar expressions for ϵ_y and ϵ_z .

In order to make use of this expression it is necessary to have some idea of the values of Δe usually encountered in experiment.

There has been considerable speculation regarding the accuracy attainable in the estimation of the intensity of single crystal reflexions, some extremely interesting work by Lonsdale (1944) suggested that crystal perfection caused extinction effects in many more low-intensity reflexions than had been previously supposed. The great divergence between observed and calculated $|F|$ values in actual structure determinations (c. 15–20 %) again suggested great experimental inaccuracies. The latter estimate of error is, in the opinion of the author, not reliable, as the calculated $|F|$ values are based on a distribution of electrons spherically symmetrical about the centre of the atoms, whereas, in fact, electrons must partake in the formation of chemical bonds and the distribution cannot be symmetrical. This is borne out by experimental work recently quoted by Robertson (1945). Another suggested source of error lies in the fact that in reducing the observations no correction is made for absorption. This again is probably not significant, since the crystals normally used in work on organic structures have low absorption coefficients and are of extremely small size (c. 0.2 mm.). Experimental evidence on these points has recently become available during a redetermination of the structure of dibenzyl by Jeffrey (1945);

this structure was originally determined by Robertson (1934, 1935) using two-dimensional methods, and a comparison of the two sets of $F(h0l)$, $F(hk0)$ and $F(0kl)$ revealed the following errors:

error	mean no. of obs.
0	45.0
1	18.5
2	4.5
3	1.0
4	0.5

These are plotted in the form of a frequency curve in figure 2. From this graph, which approaches closely to a normal error curve, it can be calculated by the usual procedure (Worthing & Jeffner 1943) that the probable error is ± 0.6 .

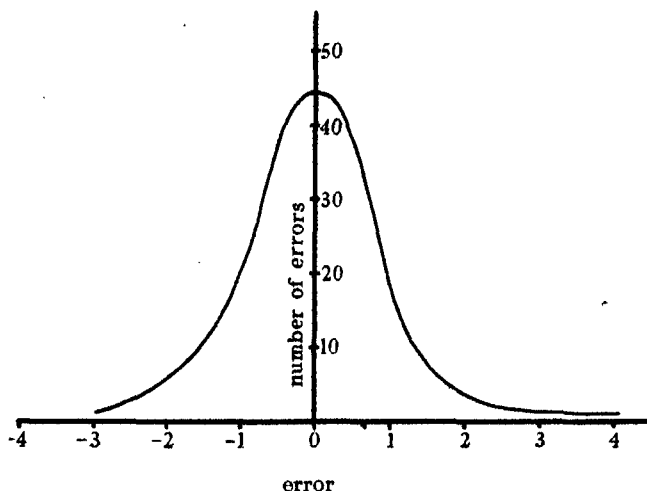


FIGURE 2. Error distribution curve.

As an example of the use of equation (5.17) it is interesting to examine the errors in co-ordinates derived from three-dimensional syntheses in a particular case. For simplicity of calculation a hypothetical structure having

$$\begin{aligned}
 [a] &= 6.156 \text{ \AA}, & a^* &= 0.25, \\
 [b] &= 12.312 \text{ \AA}, & b^* &= 0.125, \\
 [c] &= 7.695 \text{ \AA}, & c^* &= 0.20, \\
 V &= 583.2 \text{ \AA}^3,
 \end{aligned}$$

was chosen. This gives $H = 8$, $K = 16$, $L = 10$.

Substitution in (5.17) gives for a carbon atom having $N = 6$

$$\epsilon_x < 0.0026 \text{ \AA}, \quad \epsilon_y < 0.0027 \text{ \AA}, \quad \epsilon_z < 0.0027 \text{ \AA}.$$

The agreement between these values is confirmation of the validity of the approximations used. By assuming H to be large (5.17) may be written approximately

$$\epsilon_x < 90.8 \Delta e / N (\lambda p)^{\frac{1}{2}} \sqrt{V}, \quad (5.18)$$

which is symmetrical in a , b and c as suggested by the above results. The value of ϵ_x calculated from (5.18) is

$$\epsilon_x < 0.0027 \text{ \AA}$$

in very good agreement with the results of the exact calculation.

In view of this agreement, and since it is better adapted to computation than (5.17), it is suggested that equation (5.18) be used as a basis for the assessment of inaccuracies due to experiment. If the value of ' p ' used in this paper is accepted, equation (5.18) may be written

$$\epsilon < \pm 1.915 \Delta e / N \lambda^2 \sqrt{(\lambda V)}. \quad (5.19)$$

It may be contended by some workers that the probable error Δe is not constant for all observations, especially when photometric measurement is used. A reasonable assumption in this case is that the error is proportional to the index of the plane observed. Introducing this assumption into (5.12) it can be shown that the error becomes approximately

$$\epsilon < \pm 100 \Delta e / N \lambda V p^{\frac{1}{2}} \sqrt{\lambda}, \quad (5.20)$$

where Δe is, in this case, the *proportional* experimental error. A still more refined treatment would consider the experimental error to be proportional to the $|F|$ value of the plane; this, however, renders the mathematics intractable, and since, for the normal values of λ , the assumptions involved in deducing (5.19) appear to be experimentally justified, it was not considered worth while attempting an approximate solution in this case.

PART II

6. THE EFFECT OF NON-INFINITE SUMMATION

Suppose that in the unit cell of a crystal there are atoms in positions

$$(x_r, y_r, z_r) \quad (r = 1, \dots, M). \quad (6.1)$$

In accordance with (4.3) and the periodic nature of the crystal lattice the density distribution may be written

$$D(x, y, z) = \sum_{r=1}^M \left\{ \sum_{\substack{u, v, w \\ -\infty}}^{+\infty} N_r \left(\frac{p}{\pi} \right)^{\frac{1}{2}} \exp \{ -p[(x - x_r - ua)^2 + (y - y_r - vb)^2 + (z - z_r - wc)^2] \} \right\}. \quad (6.2)$$

Next, following the original concept of Bragg (1915), assume that $D(x, y, z)$ can be expressed in the form of a three-dimensional Fourier series:

$$D(x, y, z) = \sum_{h,k,l} C(h, k, l) \cos \theta + S(h, k, l) \sin \theta. \quad (6.3)$$

It then follows from the ordinary theory of Fourier transforms (Titchmarsh 1937) that

$$C(h, k, l) = \frac{1}{V} \iiint_0^{abc} D(x, y, z) \cos \theta dx dy dz \quad (6.4)$$

and

$$S(h, k, l) = \frac{1}{V} \iiint_0^{abc} D(x, y, z) \sin \theta dx dy dz. \quad (6.5)$$

Substituting the value of $D(x, y, z)$ given in (6.2) and making certain obvious transformations

$$C(h, k, l) = \frac{1}{V} \left(\frac{p}{\pi}\right)^{\frac{1}{2}} \sum_{r=1}^M N_r \iiint_{-\infty}^{\infty} \exp \{-p[(x-x_r)^2 + (y-y_r)^2 + (z-z_r)^2]\} \cos \theta dx dy dz, \quad (6.6)$$

with a similar relation for $S(h, k, l)$.

Using the well-known integral (De Hahn 1867)

$$\int_{-\infty}^{\infty} e^{-px^2} \sin q(x+\lambda) dx = \sqrt{\frac{\pi}{p}} e^{-q^2/4p} \sin q\lambda, \quad (6.7)$$

together with substitutions of the type

$$X = x - x_r, \quad (6.8)$$

(6.6) reduces to

$$C(h, k, l) = \frac{1}{V} \sum_{r=1}^M N_r \exp \left[-\frac{\pi^2}{p} \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \right] \cos \theta_r, \quad (6.9)$$

$$\text{and similarly } S(h, k, l) = \frac{1}{V} \sum_{r=1}^M N_r \exp \left[-\frac{\pi^2}{p} \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \right] \sin \theta_r, \quad (6.10)$$

or, if the series of (6.3) be written in the form given in (3.1), it is seen that

$$\left\{ \begin{aligned} |F(h, k, l)| &= V \{C^2(h, k, l) + S^2(h, k, l)\}^{\frac{1}{2}}, \\ \tan \alpha &= S(h, k, l)/C(h, k, l). \end{aligned} \right\} \quad (6.11)$$

Thus identifying 'C' and 'S' with the 'A' and 'B' parts (Lonsdale 1944) of the structure amplitude normally deduced from geometrical considerations and showing that

$$N_r \exp \left[-\frac{\pi^2}{p} \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \right] \quad (6.12)$$

is the atomic scattering factor for a distribution of type 4.2 in agreement with its value, determined from the integral (Compton & Allison 1936)

$$f = \int_0^{\infty} 4\pi r^2 d(r) \frac{\sin(2\pi \rho r/\lambda)}{2\pi \rho r/\lambda} dr, \quad (6.13)$$

where

$$\frac{\rho}{\lambda} = \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right)^{\frac{1}{2}}. \quad (6.14)$$

The identity of (6.3) and (3.1) having been demonstrated, substitution from (6.9) and (6.10) in (6.3) gives

$$D(x, y, z) = \frac{1}{V} \sum_{r=1}^M \left[N_r \sum_{h,k,l} \sum_{-\infty}^{\infty} \exp \left[-\frac{\pi^2}{p} \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \right] \cos(\theta - \theta_r) \right], \quad (6.15)$$

from which it is at once obvious that the whole summation is composed of M parts, one of which is completely in phase at each atom. Considering, for example, the atom at (x_1, y_1, z_1) , its own component in (6.15), viz.

$$\frac{N_1}{V} \sum_{h,k,l} \sum_{-\infty}^{\infty} \exp \left[-\frac{\pi^2}{p} \left(\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \right] \cos(\theta - \theta_1) \quad (6.16)$$

always has its maximum at (x_1, y_1, z_1) , however few or many terms of the series may be taken. The effect of the remaining $(M - 1)$ atomic series will be to produce distortion depending on the distances of their centres from (x_1, y_1, z_1) . Since the shape of a peak, at a distance from its maximum, will deviate slightly from that postulated in (4.2), due to finite summation, the effect of cosine waves in the position of (x_1, y_1, z_1) must also be considered.

7. DISTORTING EFFECT OF NEIGHBOURING PEAKS

Consider the effect of a peak at (x_2, y_2, z_2) on that at (x_1, y_1, z_1) the peaks being represented by equations of type (4.1). The maximum deviation will evidently be along the line of centres, and supposing the inter-peak distance to be ' s ' the resultant may be written

$$d' = N_1 e^{-pr^2} + N_2 e^{-p(r-s)^2} \quad (7.1)$$

if d' is to be a maximum $dd'/dr = 0$, whence

$$rN_1 + (r-s)N_2 e^{2prs} e^{-ps^2} = 0. \quad (7.2)$$

Assuming the maximum to occur when r is small, this gives

$$rN_1 + (r-s-2prs^2)N_2 e^{-ps^2} = 0,$$

whence

$$r = N_2 s / N_1 e^{ps^2} + 1 - 2ps^2. \quad (7.3)$$

Since the closest approach likely to be encountered in practice is about 1 Å, a good approximation to (7.3) is

$$r = \frac{N_2}{N_1} s e^{-ps^2}. \quad (7.4)$$

The values of r —the distortion from zero or true position—are given in the following table, for the case $N_1 = N_2$, calculated from the exact formula (7.3):

s (Å)	r (Å)
1.20	0.0014
1.30	0.0005
1.40	0.0001
1.50	0.0000 ₄

Hence it is seen that deviations resulting from this source are negligible.

8. UNDULATION OF BACKGROUND

The treatment given in § 6 shows that it is sufficient to consider the background undulation at a point at some distance from the maximum of a peak which may be centred on the origin. The summation in this case becomes, from (6.14) and (6.16),

$$D(x) = \frac{N}{V} \sum_{h,k,l} \sum_{\infty} \exp \left[-\frac{\pi^2 \rho^2}{p\lambda^2} \right] \cos 2\pi h \frac{x}{a} \quad (8.1)$$

along the x axis, which special case may be taken without loss of generality since the crystal axes (a, b, c) are still arbitrary. Suppose next that all planes whose ρ value is less than some specified value ρ_1 are to be included, using the concept of the reciprocal lattice and sphere already mentioned (Part I, § 5), it is seen that for $h = h_1$ (say)

$$\begin{aligned} \sum_k \sum_l \exp \left[-\frac{\pi^2 \rho^2}{p\lambda^2} \right] &\simeq \frac{2\pi}{b^*c^*} \int_{h_1 a^*}^{\rho_1} \rho \exp \left[-\frac{\pi^2 \rho^2}{p\lambda^2} \right] d\rho \\ &= \frac{p\lambda^2}{\pi b^*c^*} \left\{ \exp \left[-\frac{\pi^2 h_1^2 a^{*2}}{p\lambda^2} \right] - \exp \left[-\frac{\pi^2 \rho_1^2}{p\lambda^2} \right] \right\}, \end{aligned} \quad (8.2)$$

whence (8.1) can be written

$$D(x) \simeq \frac{Np}{\pi a} \sum_{-H}^H \exp \left[-\frac{\pi^2 h^2}{pa^2} \right] \cos 2\pi h \frac{x}{a} - \frac{Np}{\pi a} \exp \left[-\frac{\pi^2 \rho_1^2}{p\lambda^2} \right] \sum_{-H}^H \cos 2\pi h \frac{x}{a}. \quad (8.3)$$

This equation indicates that the relative lengths of the b and c axes do not effect the accuracy of the x co-ordinate. Using the well-known trigonometrical formula for the sum of the cosine series (8.3) may be written

$$D(x) \simeq \frac{Np}{\pi a} \left\{ 1 + 2 \sum_1^H \exp \left[-\frac{\pi^2 h^2}{pa^2} \right] \cos 2\pi h \frac{x}{a} \right\} - \frac{Np}{\pi a} \exp \left[-\frac{\pi^2 \rho_1^2}{p\lambda^2} \right] \frac{\sin \pi(2H+1)x/a}{\sin \pi x/a}. \quad (8.4)$$

If ρ_1 is infinite
$$D(x) = \frac{Np}{\pi a} \sum_{-H}^H \exp \left[-\frac{\pi^2 h^2}{pa^2} \right] \cos 2\pi h \frac{x}{a}, \quad (8.5)$$

which is identical with the result of expanding the distribution function

$$D(x) = \sum_{-\infty}^{\infty} N \left(\frac{p}{\pi} \right)^{\frac{1}{2}} e^{-p(x-na)^2}$$

in a one-dimensional Fourier series, thus justifying the assumptions made in obtaining equation (8.2).

The Fourier series of (8.4) is very rapidly convergent, and the background at any stage is at greatest equal to the next term to be inserted, whence the background is given approximately by the function

$$\begin{aligned} \frac{Np}{\pi a} \left\{ 2 \cos 2\pi H \frac{x}{a} - \frac{\sin \pi(2H+1)x/a}{\sin \pi x/a} \right\} \exp \left[-\frac{\pi^2 \rho_1^2}{p\lambda^2} \right] \\ = \frac{Np \sin \pi(2H-1)x/a}{\pi a \sin \pi x/a} \exp \left[-\frac{\pi^2 \rho_1^2}{p\lambda^2} \right]. \end{aligned} \quad (8.6)$$

This expression is somewhat cumbersome for combination with an atomic peak, but since peaks are not in general closer than say 1.0 Å, (8.6) may be written, replacing $\sin \pi x/a$ by a constant 'k'

$$\frac{Np}{ka\pi} \sin(2H-1)\pi \frac{x}{a} \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right]. \quad (8.7)$$

Suppose that this term has a phase shift ψ and is acting on the typical peak

$$d(x) = N_1 \left(\frac{p}{\pi}\right)^{\frac{1}{2}} e^{-px^2}.$$

The resultant is

$$d'(x) = N_1 \left(\frac{p}{\pi}\right)^{\frac{1}{2}} e^{-px^2} + \frac{Np}{ka\pi} \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right] \sin\left[2\pi(H-\frac{1}{2})\frac{x}{a} + \psi\right]. \quad (8.8)$$

The condition for a maximum is

$$-2N_1 p \left(\frac{p}{\pi}\right)^{\frac{1}{2}} x e^{-px^2} + \frac{(H-\frac{1}{2})2Np}{ka^2} \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right] \cos\left[2\pi(H-\frac{1}{2})\frac{x}{a} + \psi\right] = 0. \quad (8.9)$$

And, if the value of x satisfying this equation is to be itself a maximum, $dx/d\psi = 0$, whence

$$\sin\left[2\pi(H-\frac{1}{2})\frac{x}{a} + \psi\right] = 0,$$

or
$$\psi = n\pi - 2\pi(H-\frac{1}{2})\frac{x}{a}. \quad (8.10)$$

Substituting in (8.9) and assuming x small enough to neglect second order terms gives

$$x = \frac{N}{N_1} \frac{(H-\frac{1}{2})}{ka^2} \left(\frac{\pi}{p}\right)^{\frac{1}{2}} \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right]. \quad (8.11)$$

If H is large this may be written, to the same degree of approximation,

$$x = \frac{N}{N_1} \left(\frac{\pi}{p}\right)^{\frac{1}{2}} \frac{\rho_1}{ka\lambda} \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right]. \quad (8.12)$$

The following table contains values of the auxiliary function

$$A = \frac{1}{\lambda} \left(\frac{\pi}{p}\right)^{\frac{1}{2}} \rho_1 \exp\left[-\frac{\pi^2 \rho_1^2}{p\lambda^2}\right], \quad (8.13)$$

together with values of x , assuming $a = 6.156$ Å as in § 5, $N = N_1$, $k = 2/\pi$ (the mean values of $\sin \pi(x/a)$ when $0 < x < a$):

	A	x (Å) ($a = 6.156$ Å)
1.5	0.072	0.019
6	0.059	0.015
7	0.046	0.012
8	0.036	0.009
9	0.027	0.007
2.0	0.020	0.005
1	0.015	0.004
2	0.011	0.003

It must be realized that these figures are upper limits to the errors produced due to termination of the Fourier series at a finite number of terms. The *probability* of any of these errors is extremely small, and the more so when an assembly of atoms is considered.

9. EFFECT OF THERMAL MOTION OF ATOMS ON ACCURACY

Equation (8.12) gives immediate information as to the effect of thermal motion on accuracy, the effect of temperature on the atom itself is to produce a general spreading, that is, to decrease the value of ' p '. At the same time the greatest value of ρ_1 , which appears in conjunction with any diffracted beam, is reduced, owing to the rapid decrease of the atomic scattering factor (6.12) making observation of diffraction spots of high order impossible.

This decrease in atomic scattering factor due to decrease in ' p ' value for a thermally agitated atom is in accord with the theory advanced by Debye (1914) and Waller (1927), ' p ' being related to the quantity ' B ' defined by these authors through the equation

$$\frac{1}{p} = a(B + c),$$

where a and c are constants.

The fact that atomic peaks resulting from Fourier synthesis can be represented by (4.2) is a direct consequence of the thermal motion, if measurement of $|F|$ values was carried out at absolute zero the atomic field should have the form indicated by wave-mechanical considerations.

It can be seen that the effect of thermal motion in a structure is a decrease in the accuracy of the derived co-ordinates. Owing to the lack of knowledge of values of B for complex crystals exact quantitative expressions cannot be obtained, this, however, is not important since a method is given in the next section by which the errors in any particular structure can be determined.

10. ASSESSMENT OF ACCURACY IN ANY GIVEN STRUCTURE DETERMINATION

Equation (8.12) gives only an upper limit to the error due to finite summation in the di-atomic case. Since the positions of atoms form a predetermined system statistical laws cannot be applied to give most probable errors. However, equation (6.15) shows how the error, due to finite summation, may be determined for any atom in a particular structure. The co-ordinates used for calculating F values are those involved in the θ , of (6.15). If, therefore, when a structure determination is complete, an additional synthesis is carried out using the *calculated* F values, the deviations of the resulting maxima from the positions used to calculate the F values give the errors in the resulting co-ordinates. These errors, being in general small, can be applied with reversed signs to the co-ordinates derived from the $F(\text{obs.})$ synthesis to give more accurate values of the latter and the inaccuracies due to termination of the summation almost completely removed.

A trial of this method of assessing accuracy has been made on the compound geranylamine hydrochloride already quoted (Jeffrey 1945) and deviations of the same order (0.02 Å) as predicted in § 8 have been found in the co-ordinates due to finite summation. It follows that the experimental errors (c. 0.002 Å in this case) are negligible in comparison with the finite summation errors.

Two courses are thus open, the derived co-ordinates may be preserved and assumed to be in error by the quantities found in the above 'blank' synthesis, or, alternatively, the co-ordinates may be corrected as indicated and the errors then taken as double those due to experiment calculated on the basis of (5.18).

11. CONCLUSIONS

As a result of the foregoing treatment the following general statement of the accuracy of three-dimensional Fourier syntheses can be made:

(1) Expressions for probable errors in co-ordinate have been given (5.17–5.20) in terms of the experimental errors usually encountered. The expressions have been applied to a particular case and show that for a carbon atom the error is approximately ± 0.003 Å.

(2) A method for determining the errors due to non-infinite limits of summation (§ 10) has been devised and indications given of the method of applying these results to the correction of the co-ordinates deduced in any given structure. In the special case of a structure containing two atoms an upper limit of between ± 0.02 and ± 0.005 Å to the distortion is deduced, depending on the maximum order of reflexion observed. In a typical polyatomic structure these errors have been found experimentally to be of the order ± 0.02 Å, so that the errors considered in (1) are of the second order.

(3) The effect of thermal agitation is shown to be a considerable decrease in accuracy, its exact amount being governed by the criteria laid down in (2).

The author wishes to record his indebtedness to Professor G. N. Watson, F.R.S., for a most helpful discussion on the subject of the second part of this paper. This work forms part of the programme of fundamental research undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

- Bernal 1926 *Proc. Roy. Soc. A*, **113**, 117–141.
Booth 1945 *Trans. Faraday Soc.* **41**, 434.
Booth & Llewellyn Work at present unpublished under security regulations.
Bragg, W. H. 1915 *Phil. Trans.* **215**, 253.
Compton & Allison 1936 *X-rays in theory and experiment*. London: Macmillan.
Costain 1941 Ph.D. Thesis, University of Birmingham.
Cox & Jeffrey 1942 *Trans. Faraday Soc.* **37**, 241–247.
Debye 1914 *Ann. Phys.*, Paris, **43**, 49.

- De Hahn 1867 *Tables d'intégrales définies*. New York: Steckert.
Ewald 1921 *Z. Kristallogr. A*, **56**, 148-150.
Hartree 1928 *Proc. Camb. Phil. Soc.* **24**, 89.
Hartree 1928 *Proc. Camb. Phil. Soc.* **24**, 111.
Jeffrey Private communication.
Jeffrey 1945 *Proc. Roy. Soc. A*, **183**, 388.
Lonsdale 1936 *Simplified structure factor tables*. London: Boll.
Lonsdale 1944 *Nature*, **153**, 22 and 433.
Preston 1944 *J. Sci. Instrum.* **21**, 205-213.
Robertson 1934 *Proc. Roy. Soc. A*, **146**, 473.
Robertson 1935 *Proc. Roy. Soc. A*, **150**, 348.
Robertson 1945 Tilden Lecture, Chemical Society.
Titchmarsh 1937 *Introduction to the theory of Fourier integrals*. Oxford University Press.
Waller 1927 *Z. Phys.* **83**, 153.
Worthing & Jeffner 1943 *Treatment of experimental data*. New York: Wiley.
-

The mechanism of the catalytic oxidation of ethylene

I. Experiments using a flow system

By G. H. TWIGG, PH.D. *Department of Colloid Science, Cambridge*

(Communicated by E. K. Rideal, F.R.S.—Received 31 July 1945)

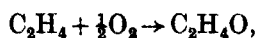
Experiments have been made using a flow system to determine the mechanism of the catalytic oxidation of ethylene on a silver catalyst. The effects of time of contact of the gases with the catalyst, gas concentration, and temperature have been investigated. The products of reaction are ethylene oxide, and carbon dioxide and water. There appear to be two processes whereby the carbon dioxide is formed: (1) by direct oxidation of the ethylene not via ethylene oxide, and (2) by the further oxidation of the ethylene oxide.

The isomerization of ethylene oxide to acetaldehyde by the catalyst in the absence of any oxygen has also been examined. By comparison with the oxidation of ethylene oxide, it has been shown that this latter reaction proceeds to a large extent, and possibly entirely, through a preliminary isomerization of the ethylene oxide to acetaldehyde. The rate of oxidation of acetaldehyde is extremely rapid and no trace of acetaldehyde is found during the oxidation of ethylene or of ethylene oxide. Ethylene oxide forms on the catalyst an involatile deposit, which is oxidized away by oxygen, so that during oxidation reactions the quantity of it on the catalyst is kept low.

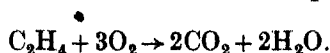
The kinetics of the oxidation of ethylene, i.e. rate of reaction proportional to the oxygen concentration and slightly dependent on the ethylene pressure, are consistent with the view that ethylene reacts with oxygen adsorbed on the catalyst and that the slowest step in the whole series of reactions is the rate of adsorption of the oxygen. An energy of activation of about 27 kcal. was found for the production of ethylene oxide, and slightly less for the production of carbon dioxide and consumption of oxygen.

In the homogeneous reaction between ethylene and oxygen at temperatures in the region of 300-600° C, it has been found (Lehner 1931; Bone, Haffner & Rance 1933) that a number of products are formed, including ethylene oxide, aldehydes,

formic acid, oxides of carbon and polymers of ethylene. On the other hand, when the reaction is carried out with the aid of a silver catalyst, the products are much simpler, being ethylene oxide formed according to the equation



and carbon dioxide and water obtained from the complete combustion



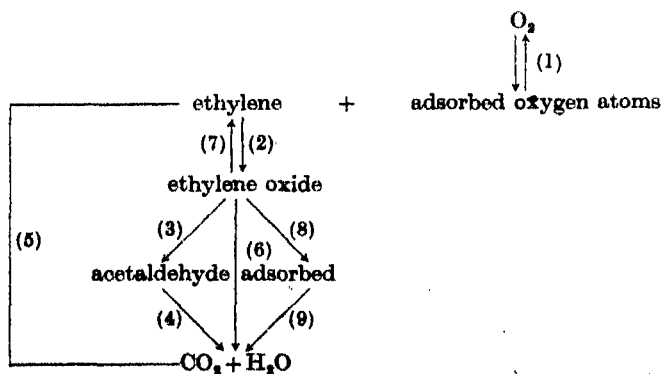
The purpose of the present work has been to investigate the mechanism of this catalytic reaction.

The fundamental role of the catalyst is believed to be the adsorption of oxygen as atoms; reaction then occurs between ethylene (and other substances) and the adsorbed oxygen atoms. The reactions occurring have proved to be very complex, and a variety of lines of attack on the problem have been followed. In the first place, experiments were made in a flow system, in which the gases were passed over the catalyst and the products analysed. The advantage of this method is that a steady state of adsorption on the catalyst is reached. The disadvantages are that flow rates and temperature are not susceptible of accurate measurement.

In the second part of the paper, reactions will be considered between various organic substances and the adsorbed oxygen layer, and between these substances and the deoxygenated catalyst. The method used here indicates most of the reactions occurring, but not their relative importance; it has to be considered in conjunction with the results obtained by other methods.

Thirdly, reactions in a static system between ethylene, etc., and oxygen were examined, from which the final picture was constructed. Information was also obtained from a peculiar effect which was observed to take place on the electrical resistance of the particular catalyst used here and which was caused by the chemisorption of oxygen.

The scheme of the reactions found to be taking place during the oxidation of ethylene is given in the accompanying diagram. Reactions 6-9 are believed to be of less importance than the others:



THE CATALYST

The catalyst consisted of 'Versil' glass-wool, diameter of fibres $c. 10^{-2}$ mm., plated with a thin film of silver by reduction of an ammoniacal silver solution with sucrose, according to the method of Common (1890). This catalyst has the advantages of a large surface area for a given weight of silver, and also good thermal conductivity to minimize local heating. This latter is important, since the oxidation to ethylene oxide, and, to a greater degree, the complete combustion to carbon dioxide and water, are strongly exothermic. The effect of packing on the thermal conductivity was estimated from the behaviour of the electrical resistance. The catalyst was loosely packed into a glass tube of 20 mm. diameter, and confined between two closely fitting copper rods, one of which was fixed and the other movable. The catalyst was compressed in stages and the resistance measured for varying length of catalyst in the tube. Over the complete range from 3 to 10^{-2} Ω , the resistance was proportional to the cube of the length of the catalyst. Hence, the specific conductivity of the catalyst is proportional to the square of its density. Since the amount of reaction and so the heat produced is proportional to the first power of the density, and assuming proportionality between the electrical and thermal conductivities, it thus appears advantageous to use the catalyst tightly compressed to ensure uniform temperature distribution.

The thickness of the silver plating, as estimated by determining the percentage silver in the catalyst, and also by measuring the electrical resistance of individual fibres, was about 10^{-4} mm. It was found necessary to plate the catalyst at least twice, as single-plated samples deteriorated with use. This deterioration was found to be due to the aggregation of the thin silver film into isolated globules, thus decreasing the area of the surface, and breaking its continuity and thereby decreasing the thermal conductivity; measurements of the electrical resistance confirmed this view.

APPARATUS AND METHOD

The apparatus used in these experiments is shown in figure 1. Compressed air and cylinder ethylene after drying were passed through flowmeters of the adjustable type (Farkas & Melville 1939). After mixing, the gases passed through the electrically heated pyrex reaction vessel. The reaction vessel had a central thermocouple pocket (*Th*). The gases, after reaction, were collected for analysis in the traps S_1 , S_2 and S_3 . The trap S_1 was cooled to -20°C by an ice-salt bath; here most of the water formed in the reaction was removed. Trap S_2 was immersed in an acetone-solid CO_2 bath at -80°C to collect the major part of the ethylene oxide. In the last trap, which was cooled in liquid air, the remainder of the ethylene oxide, together with the carbon dioxide and residual ethylene, were collected. By suitable manipulation of the taps, the contents of each trap could be separately pumped over and collected in S_4 , separated from air, and analysed. The ethylene oxide was condensed in the graduated tube *R* and the volume of liquid measured; the density of the liquid was taken as

0.89 g./c.c. at 5° C (Timmermans & Hennaut-Roland 1937). The small quantity of ethylene oxide in S_3 was separated by a crude fractionation in S_4 from the ethylene and carbon dioxide which were transferred to the bulb B of known volume; from the pressure in B the sum of the volumes of ethylene and carbon dioxide was then known. The separate quantities of these were found from an analysis of the carbon dioxide in the mixture obtained by transferring part of the contents of B to an Orsat apparatus. Water could be estimated by condensing in a detachable side tube and weighing, but was usually taken as equivalent to the carbon dioxide according to the equation

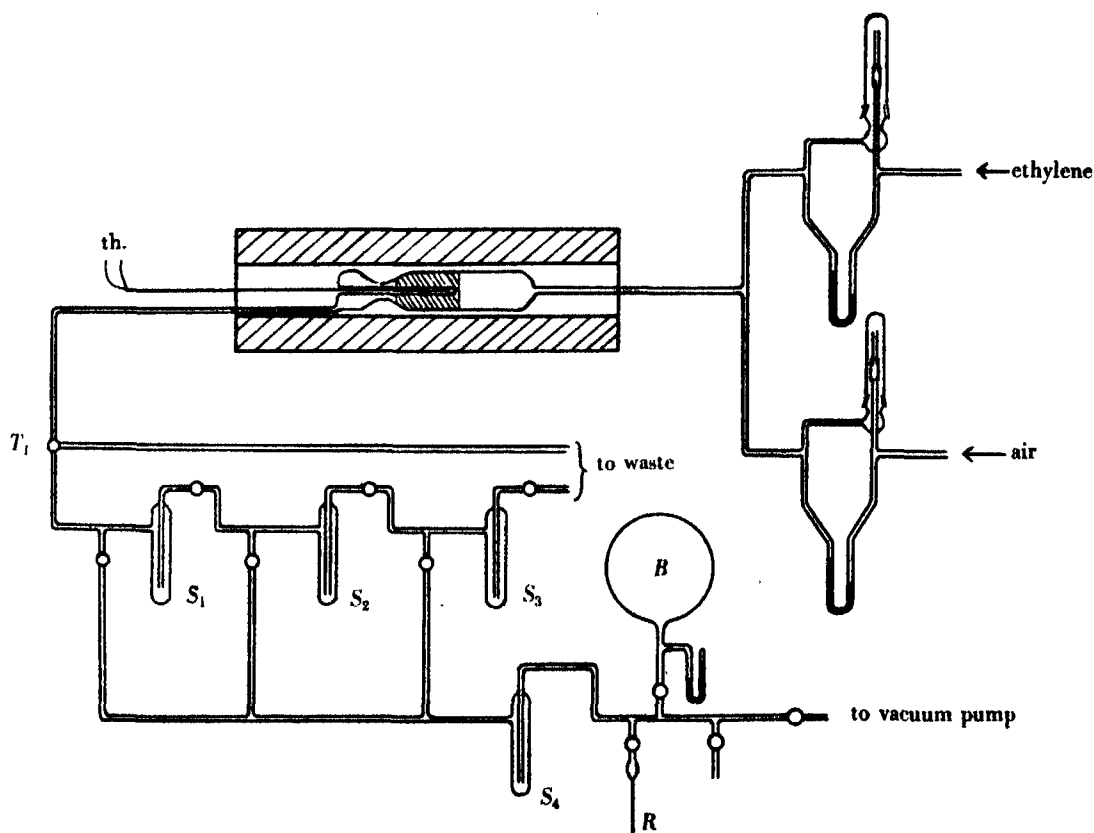
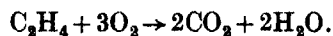


FIGURE 1. Apparatus.

The method of analysis was checked by using artificial mixtures; the ethylene oxide estimation was always slightly low owing to inefficient separation of the small quantity collected along with the carbon dioxide and ethylene.

The catalyst and experimental arrangement described were found to produce satisfactory yields of ethylene oxide. For example, at 255° C, with flow rates of 5 c.c. ethylene and 50 c.c. air per minute, 13.6 % of the ethylene was converted to ethylene

oxide, and 4.0 % to carbon dioxide, a yield of 77 % on the ethylene reacted. The temperature of reaction was generally between 240 and 350° C, the lower temperatures favouring the production of ethylene oxide relative to the carbon dioxide. At the higher temperatures, owing to the strongly exothermic nature of the complete combustion ($\text{C}_2\text{H}_4 + 3\text{O}_2 \rightarrow 2\text{CO}_2 + 2\text{H}_2\text{O} + 333 \text{ kcal.}$), temperature control was difficult.

MECHANISM OF REACTION

(1) *Ethylene-oxygen experiments*

Using the present catalyst and experimental conditions, the only intermediate which has been detected between ethylene and the end-products of the oxidation, carbon dioxide and water, is ethylene oxide; any other products, such as are found in the homogeneous reaction, are too short-lived to be present in detectable amounts. In elucidating the mechanism of the oxidation, the first question to be decided is whether the carbon dioxide and water arise from the further oxidation of the ethylene oxide or whether they are formed from the ethylene by another route involving such short-lived intermediates. The effect of contact time might be expected to throw some light on this point.

If the carbon dioxide and water are formed via ethylene oxide, then the relationship between the percentage of ingoing ethylene converted to (a) ethylene oxide, and to (b) carbon dioxide, and the time of contact should have the form shown in figure 2 (full curves); there should be an induction period at the beginning of the carbon dioxide curve while the ethylene oxide concentration builds up, so that the ratio carbon dioxide to ethylene oxide tends to zero as the time tends to zero, and there should be a maximum in the ethylene oxide curve.

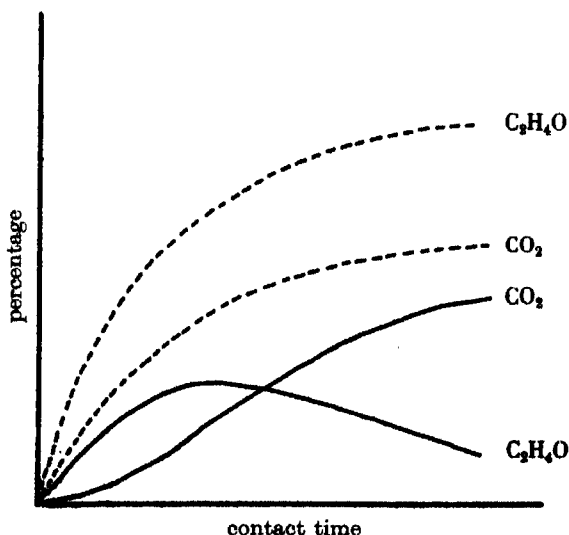


FIGURE 2. Reaction curves for alternative mechanisms of oxidation.

If, on the other hand, the carbon dioxide does not arise from the ethylene oxide, it must be formed through a series of short-lived intermediates, which do not build up in the system; thus there should be no detectable induction period in the carbon dioxide curve, the ethylene oxide curve should rise continuously, and the ratio of carbon dioxide to ethylene oxide should remain substantially constant with time. The curves should have the form shown in figure 2 (dotted lines). For convenience, this alternative route for the production of carbon dioxide and water has been termed the 'direct oxidation of ethylene', though this is not to imply that intermediates do not exist which could not, under suitable conditions, be isolated.

The experimental results obtained are given in table 1. Catalysts of three different activities were used. In the first two groups of experiments at low conversions, there is no sign of an induction period; the production of carbon dioxide by a route not involving ethylene oxide is thus indicated. On the other hand, the third group show that a maximum in the ethylene oxide is being reached while the carbon dioxide continues to increase, as would be required by the hypothesis that the carbon dioxide arises from the ethylene oxide. These experiments thus suggest that both processes operate and the carbon dioxide arises directly from the ethylene and also from the further oxidation of the ethylene oxide.

The effect of ethylene concentration on the rate of reaction can be seen from table 2. The contact time was kept constant, and the ethylene concentration varied over a four-fold range.

TABLE 1. EFFECT OF CONTACT TIME

experiment number	catalyst	T° C	contact time (min.)	% conversion to	
				EtO	CO ₂
042	8	340	0.25	3.6	2.5
043	8	340	0.50	4.2	4.4
044	8	338	1.0	8.8	8.8
051	10	320	0.25	3.8	3.2
052	10	328	1.0	13.7	14.0
92	24D	305	0.23	34.0	18.8
90	24D	304	0.43	39.5	24.7
94	24D	305	0.90	41.4	31.4

TABLE 2. EFFECT OF ETHYLENE CONCENTRATION

experiment number	T° C	partial pressures (mm.)			
		reactants		products	
		Et	O ₂	EtO	CO ₂
072	329	217	109	15.9	26.1
073	328	109	131	13.3	—
074	328	54	141	9.1	15.6
075	330	54	141	9.0	25.8
076	331	109	131	12.0	32.8
077	330	217	109	16.1	27.0

TABLE 3. EFFECT OF OXYGEN CONCENTRATION

experiment number	T° C	partial pressures (mm.)			
		reactants		products	
		Et	O ₂	EtO	CO ₂
076	331	109	131	12.0	32.8
081	329	109	65	8.0	13.0
082	328	109	33	3.5	8.5

It is difficult to give these results greater precision. The range of flow rates available is limited, and the analyses cannot be made to a high degree of accuracy, especially at small conversions. The chief difficulty, however, is with regard to the temperature; despite the good thermal conductivity of the catalyst, the true temperature of the catalyst may be above that recorded on the thermocouple, and hot spots may develop; this seems to happen when the conversion to carbon dioxide is high, and the reproducibility of the experiments is then not very great.

Despite these limitations, however, it can be seen from tables 2 and 3 that the rate of oxidation both to ethylene oxide and to carbon dioxide is only slightly dependent on the ethylene concentration, and proportional approximately to the first power of the oxygen concentration. This might suggest that the ethylene was moderately strongly adsorbed on the catalyst, and the oxygen only slightly adsorbed. However, it is known from the work of Benton & Drake (1934) that oxygen is strongly adsorbed on the surface of silver at temperatures up to 300° C. Furthermore, experiments made later using the resistance-drop effect (Twigg, 1946) show that oxygen is adsorbed on the catalyst in the reaction between ethylene and oxygen and that increase of ethylene concentration lowers this surface oxygen concentration. Thus the kinetics of reaction found here are to be accounted for by the fact that while the true order of reaction with respect to ethylene is unity, the apparent order is lower through the surface concentration of adsorbed oxygen being inversely related to the ethylene concentration.

The effect of temperature on the rates of oxidation was determined. Ethylene (50 c.c./min.) and air (500 c.c./min.) were passed over the catalyst at temperatures between 190 and 360° C. Table 4 shows the percentages of ethylene oxidized to ethylene oxide and to carbon dioxide, and of oxygen consumed. To allow for the effect of oxygen concentration, these figures were converted by dividing them by the average percentage of oxygen present during reaction. The values k_1 , k_2 and k_3 thus obtained are related to the reaction velocity constants. This method of approximation is only valid at low conversions as can be seen from figure 3, where $\log k$ is plotted against the reciprocal of the absolute temperature for the ethylene oxide production, and for the oxygen consumption. The points for the ethylene oxide fall off markedly from the straight line at high conversions owing to the further oxidation of the ethylene oxide; this is seen again in experiment 49. The points for oxygen consumption fall off slightly at the higher temperatures. The curve for CO₂ produc-

tion is closely parallel to that for oxygen consumption. The energy of activation deduced from the straight-line portion of the ethylene oxide curve is 27 kcal., while that for oxygen consumption is somewhat less (c. 23 kcal.).

TABLE 4. EFFECT OF TEMPERATURE

experiment number	$T^{\circ}\text{C}$	% conversion to		% oxygen consumed	EtO k_1	CO_2 k_2	O_2 k_3
		EtO	CO_2				
56	193	1.32	1.20	2.1	1.32	1.20	2.1
55	216	4.8	2.8	5.4	4.9	2.8	5.5
52	224.5	9.2	4.8	9.5	9.6	5.1	10.0
54	235	13.6	5.8	12.1	14.4	6.2	12.9
57	235	14.0	6.4	13.1	15.0	6.8	14.0
58	235	14.0	8.6	16.2	15.2	9.4	17.6
53	243.5	15.4	9.4	18.0	17.0	10.4	19.8
50	268	23.8	22.2	39.3	29.6	27.6	48.9
51	302	30.4	40.4	68.2	46.0	61.4	103.5
49	360	2.4	64.4	97.2	—	—	—

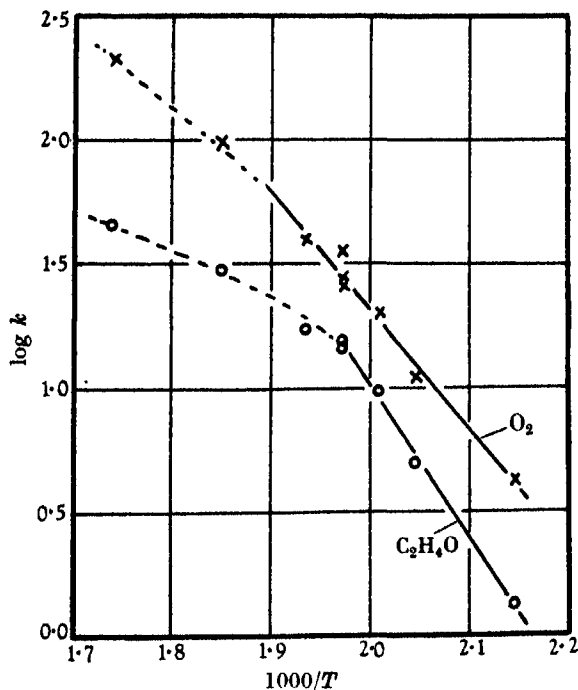


FIGURE 3. Effect of temperature on rate of oxidation. \circ Production of ethylene oxide. \times disappearance of oxygen.

(2) Ethylene oxide-oxygen experiments

Ethylene oxide was introduced into the gas stream by means of the carburettor (figure 4) which consists of a long tube containing 5–10 c.c. of liquid ethylene oxide through which air is bubbled. The upper part of the tube is surrounded by a Dewar

vessel containing ice, and the lower part is immersed in a bath of water at 12° C. The air stream picks up ethylene oxide excess of which is condensed out at 0° C. The issuing mixture thus contains ethylene oxide at the partial vapour pressure (490 mm.) at 0° C. The main air stream passes through a flowmeter to the reaction vessel as in figure 1; a subsidiary stream goes via another flowmeter to the carburettor and then joins the main stream. The capillary of this second flowmeter was chosen so that the issuing mixture contained 9 % of ethylene oxide; the accuracy of the arrangement was checked by analysis. The ethylene oxide was made in the present apparatus. After storage in sealed glass tubes for some time, some conversion to acetaldehyde occurred, from which it was purified by standing over hydroxylamine sulphate for 24 hr. or more, followed by distillation.

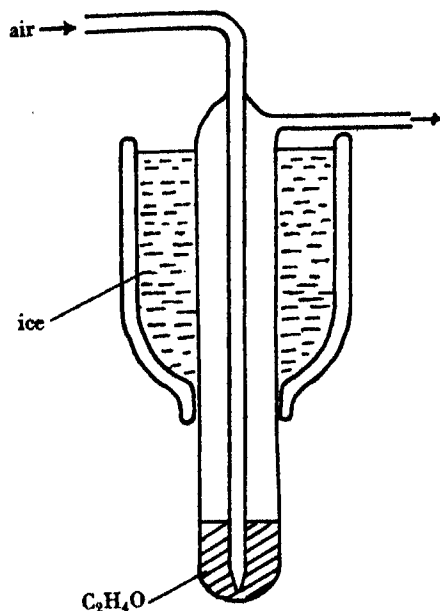


FIGURE 4. Saturator for ethylene oxide-air mixtures.

TABLE 5. $T = 323^{\circ}\text{C}$

experiment number	contact time (min.)	flow rates (c.c./min.)		% EtO oxidized
		air	EtO	
99	0.215	211	21.7	14.8
96	0.234	195	20.1	14.9
101	0.42	98.5	10.0	27.0
98	0.73	63	5.6	27.8
102	1.06	42.7	4.52	42.4
100	1.83	25	2.45	47.3

It was found that ethylene oxide was readily oxidized to carbon dioxide and water. In table 5 some experiments on the effect of contact time are listed. It will be seen that with flow rates of 100 c.c. air and 10 c.c. ethylene oxide per minute,

about 25 % of the ethylene oxide was oxidized. In a comparable experiment with ethylene and air (air 104 c.c./min.; ethylene 10.7 c.c./min.) 41.2 % of the ethylene was transformed into ethylene oxide, and 25.8 % was oxidized to carbon dioxide.

From these figures it may be concluded that part, at least, of the carbon dioxide produced in the reaction between ethylene and oxygen arises from the further oxidation of the ethylene oxide. On the other hand, since the rate of oxidation of ethylene oxide is less than that of ethylene, it seems probable that part of the carbon dioxide in the latter oxidation is produced directly from ethylene without the formation of ethylene oxide.

(3) *Isomerization of ethylene oxide to acetaldehyde*

It was suspected that the oxidation of ethylene oxide might take place through, first, isomerization to acetaldehyde which would then be rapidly oxidized. Normally in the reaction between ethylene and air, no acetaldehyde can be detected in the products; but in one experiment where the ethylene concentration was increased to 30 %, the ethylene oxide produced smelt strongly of acetaldehyde; there was insufficient oxygen to oxidize the latter as the conversion was high.

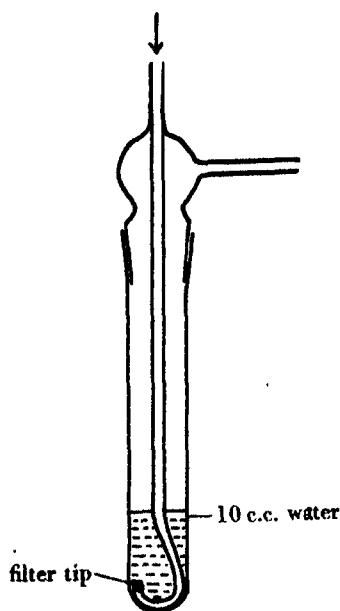


FIGURE 5. Bubbler for acetaldehyde estimations.

The apparatus used was the same as in the experiments on the oxidation of ethylene oxide. The air stream was replaced by one of nitrogen, and the gas going into the catalyst contained about 9 % of ethylene oxide. Analysis was effected by passing the reacted gases through a bubbler (figure 5) in which was placed 10 c.c. of distilled water. In all experiments the rate of gas flow was maintained approximately

constant at 130 c.c./min., and the gases were passed through the bubbler for exactly 5 min. Five c.c. of the solution so formed was then made up to a standard volume, usually 50 c.c., with distilled water, and this solution was analysed colorimetrically by means of Schiff's reagent, as described elsewhere. The efficiency of the extraction of the aldehyde from the gas stream was checked by carrying out comparable runs at room temperature with acetaldehyde in place of the ethylene oxide in the carburettor; it was found to be about 87 %. The bubbler could be by-passed and the gases collected and analysed as described before. It was verified that ethylene oxide had no effect on the analysis of the acetaldehyde.

The experimental results in table 6 show that ethylene oxide does in fact readily isomerize to acetaldehyde. But these figures also show a steady decrease in the isomerizing power of the catalyst; the first five experiments were carried out in sequence without the catalyst being in contact with any other gas than nitrogen and ethylene oxide. Between experiments 142 and 145, air was passed over the catalyst and experiment 145 was made shortly after starting the stream of nitrogen and ethylene oxide; the high value of 12.2 % was then found for the conversion. Another experiment (147) confirmed this phenomenon of deactivation. Air was passed over the catalyst at a temperature of 323° C for 1 hr. Then a mixture of nitrogen (130 c.c./min.) and ethylene oxide (12 c.c./min.) was passed for 3½ hr. Three minutes after the ethylene oxide flow had started, the gases were analysed, and it was found that 11.1 % of the ethylene oxide was being converted to acetaldehyde. Three hours later the conversion had dropped to 2.3 %.

TABLE 6. $T = 323^{\circ}\text{C}$

experiment number	flow rates (c.c./min.)		% conversion to aldehyde
	N ₂	EtO	
122	120	12.5	8.2
123	124	12.7	3.1
124	123	11.8	3.2
141	118	11.0	2.7
142	117	11.1	2.2
145	117	11.4	12.2

The decrease in activity of the catalyst is due to an organic deposit forming on the catalyst. When air was passed over a deactivated catalyst, a slight rise in temperature was noted. After experiment 147 the gas flow was stopped; the catalyst was cooled to below 100° C and swept out with nitrogen for half an hour. The catalyst temperature was then raised to 326° C and nitrogen passed at the rate of 120 c.c./min. for 25 min. The gases from the catalyst were passed through a liquid air trap, in which were collected 9 c.c. of CO₂ and about 4 c.c. of water (gaseous). A stream of air was substituted for the nitrogen, and after another 25 min. an extra 1 c.c. of CO₂ and 1 c.c. of water were collected. No other product was found during the passage of the nitrogen. It seems probable that the carbon dioxide and water were formed

through the oxidation of the organic deposit by the small quantity of oxygen present in the cylinder nitrogen.

The exact nature of this deposit is not certain; it contains carbon, hydrogen and possibly oxygen, and may be a polymer of ethylene oxide. From the amount of carbon dioxide formed, the quantity of deposit on the catalyst corresponds approximately to a monomolecular layer. Experiments with acetaldehyde showed that it, too, could form a deposit on the catalyst, though the rate of formation was much less than with ethylene oxide. During the reaction between ethylene oxide and oxygen, the deposit will be oxidized off the catalyst as fast as it is formed and the surface will be largely free of it.

The possibility that the catalyst support material, glass, might be causing the isomerization was investigated. At the flow rates used above and with a quantity of glass-wool similar to that used in making the catalyst, 0.21 % of the ethylene oxide was isomerized to acetaldehyde at 323° C; even at 370° C the conversion was only 0.55 %. This possibility can thus be excluded.

Finally, an experiment was carried out in which air (123 c.c./min.) and acetaldehyde (5.1 c.c./min.) were passed over the catalyst at 322° C. The oxidation of the acetaldehyde was complete, and no trace (< 0.01 %) could be found in the issuing gases.

These experiments thus demonstrate that the silver catalyst can bring about the isomerization of ethylene oxide to acetaldehyde, and that when oxygen is present, can oxidize the acetaldehyde with extreme rapidity. The question then arises whether this is the normal course of the oxidation of ethylene oxide. In experiment 143, a mixture of air (122 c.c./min.) and ethylene oxide (12.4 c.c./min.) was passed over the catalyst at 323° C, and 28.0 % of the ethylene oxide was oxidized to carbon dioxide. (The agreement between this figure and those of table 5 shows that the catalyst is not permanently affected by the organic deposit.) Under similar conditions in the absence of any oxygen, the maximum amount found to be isomerized was 12.2 %. Since, even in this case, the catalyst will have been partially deactivated, the figure for the isomerization on the clean silver must be in excess of this. Another factor to be considered is the true temperature of the catalyst. The oxidation of ethylene oxide is highly exothermic ($-\Delta H = 288$ kcal.), while the isomerization to acetaldehyde is much less so ($-\Delta H = 27$ kcal.). Thus for a given thermocouple temperature, the true temperature of the catalyst during oxidation may be somewhat greater than that during isomerization. In comparing the two reactions, therefore, the figure of 12.2 % may, on this account also, have to be raised. It may be concluded then that in the reaction between ethylene oxide and oxygen half of the ethylene oxide oxidized, and possibly even all of it, goes through a preliminary isomerization to acetaldehyde, the subsequent oxidation of which is extremely rapid.

GENERAL CONCLUSIONS

The flow system, despite the difficulties of accurate measurement, has provided a certain amount of information about the mechanism of oxidation. When ethylene reacts with oxygen on a silver catalyst, part of it is converted into ethylene oxide. This in turn can be further oxidized to carbon dioxide and water, a reaction which appears to involve the prior isomerization of the ethylene oxide to acetaldehyde. The acetaldehyde thus formed is oxidized to carbon dioxide with extreme rapidity, and no acetaldehyde was ever found during the oxidation either of ethylene or of ethylene oxide when sufficient oxygen was present. This chain of reactions accounts for part of the carbon dioxide formed during the oxidation of ethylene. Another, though small, fraction of the carbon dioxide will be formed by the oxidation of the involatile deposit which is formed on the catalyst by the ethylene oxide. But there are indications that carbon dioxide may also arise directly from the ethylene, without passing through ethylene oxide. In the reaction between ethylene and oxygen, the ratio of carbon dioxide to ethylene oxide does not tend to zero when the amount of reaction decreases, as it should do if the carbon dioxide is being formed only from the ethylene oxide. And again, a comparison of the rates of oxidation of ethylene and of ethylene oxide suggest that the latter reaction is not fast enough to account for all the carbon dioxide. These experiments thus indicate that the carbon dioxide arises not only from the ethylene oxide, but also directly from the ethylene.

REFERENCES

- Benton & Drake 1934 *J. Amer. Chem. Soc.* **56**, 255.
Bone, Haffner & Rance 1933 *Proc. Roy. Soc. A*, **143**, 16.
Common 1890 *Proc. Roy. Instn.*, p. 171.
Farkas & Melville 1939 *Experimental methods in gas reactions*, p. 102. London: Macmillan and Co.
Lehner 1931 *J. Amer. Chem. Soc.* **53**, 3737.
Timmermans & Hennaut-Roland 1937 *J. Chim. Phys.* **34**, 693.
Twigg 1946 *Trans. Farad. Soc.*, (in the Press).

The mechanism of the catalytic oxidation of ethylene

II. Reactions between ethylene, etc. and chemisorbed oxygen monolayers

By G. H. TWIGG, PH.D. *Department of Colloid Science, Cambridge*

(Communicated by E. K. Rideal, F.R.S.—Received 31 July 1945)

Experiments have been carried out at temperatures of 263° C and higher between oxygen adsorbed as atoms on the silver catalyst, and ethylene, ethylene oxide and acetaldehyde.

The course of reaction was followed by measuring the change in pressure, and analyses of the products were made by micro-fractionation of the gases at low temperatures.

In the reaction of ethylene with an oxygen-covered catalyst, the absence of an induction period in the pressure-time curve showed that oxidation of ethylene to carbon dioxide and water by a route not through ethylene oxide is possible. The reaction of acetaldehyde with the oxygenated catalyst was too fast to measure.

The reactions of ethylene oxide were found to be complex, and reaction occurred both with the oxygenated and the clean catalyst. On a clean catalyst, ethylene oxide was simultaneously isomerized to acetaldehyde and converted back to ethylene and adsorbed oxygen; the acetaldehyde and adsorbed oxygen then reacted to form carbon dioxide and water.

Both ethylene oxide and acetaldehyde, but not ethylene, were adsorbed with decomposition to form a non-volatile layer on the catalyst. This was composed of carbon, hydrogen and possibly oxygen, combined in indefinite and varying proportions.

The kinetics of the reaction between ethylene and the adsorbed oxygen layer were measured. Throughout the course of any one reaction, the rate of oxidation to carbon dioxide was proportional to the square of the concentration of adsorbed oxygen, but the velocity constant depended on the initial concentration. The apparent energy of activation was 10 kcal.

It is thought that when ethylene reacts with a single adsorbed oxygen atom, ethylene oxide is produced, and that with a pair of adsorbed oxygen atoms, intermediates such as formaldehyde are produced which react rapidly to form carbon dioxide and water.

The silver-plated glass-wool catalyst used in this work has a highly extended surface, large enough to enable quantities of gas adsorbed on the surface to be measured easily in a vacuum apparatus by means of a simple mercury manometer. When such an apparatus was set up, it was found that oxygen was chemisorbed on the silver surface in the atomic state, and that a number of organic substances, and in particular, ethylene, ethylene oxide and acetaldehyde, reacted with this adsorbed oxygen and removed it from the surface.

A study of these reactions has furnished further information about the catalytic oxidation. The evidence points to the process of reaction between gaseous oxygen and ethylene being, first, adsorption of the oxygen as atoms, followed by reaction between the various organic substances and this adsorbed oxygen. The importance of the technique used here is that it enables us to examine these reactions in detail and to disentangle one from the other. On the other hand, the method must be used in conjunction with the other methods, as the reactions disclosed here may not be those mainly concerned when gaseous oxygen is present; for example, this technique would exclude homogeneous reactions and those in which ethylene was adsorbed.

Also, there is the possibility that the major part of the normal reaction between gaseous oxygen and ethylene may take place on a small fraction of the surface—on 'active spots'—whereas the technique used here gives information about the surface as a whole. Despite these possibilities, it is thought that, with the exception of the reversion of ethylene oxide to ethylene and adsorbed oxygen, all the reactions found here play an important part in the normal oxidation; and, by a comparison of velocity constants, it seems that the major part at least of the surface is engaged in the oxidation.

The reactions were carried out by leaving the catalyst in oxygen at the working temperature for some time in order to let the adsorbed layer form, adsorption and desorption being slow activated processes. The gaseous oxygen was pumped out of the reaction vessel and the gas under examination was admitted. The oxygen with which this reacted was in the form of adsorbed atoms; evidence for this will be presented elsewhere but that it is adsorbed and not dissolved is shown by the fact that the amount of oxygen taken up by the catalyst at a pressure of 100 mm. and a temperature of 263° C is about twenty times as much as that calculated for dissolved oxygen under similar conditions from the figures of Steacie & Johnson (1926).

APPARATUS AND METHOD

A diagram of the apparatus is shown in figure 1. Ten grams of the silver-plated glass-wool catalyst were contained in the reaction vessel *C* mounted in an electrically heated furnace; the pressure in the reaction vessel was measured on the mercury manometer M_1 . In order to prevent water vapour produced in the reaction diffusing to the colder parts of the apparatus, the tubing between the reaction vessel and the tap T_2 was made as narrow as possible (2 mm.) consistent with fairly rapid pumping of the reaction vessel; the tubing was also wound with a heating coil. The rate of pumping was such that the reaction vessel was evacuated from a pressure of $\frac{1}{2}$ atm. to 0.03 mm. pressure in 1 min. The pumping system consisted of a mercury diffusion pump backed by a Hyvac pump; the whole apparatus could be evacuated down to a pressure of 10^{-6} mm. or less.

The ethylene used was taken from a cylinder and purified by repeated distillation at liquid-air temperature, the middle third being collected and stored. The ethylene oxide was prepared by oxidation of ethylene in the apparatus described in Part I. Oxygen was generated as required by heating potassium permanganate in the tube *P*. Impurities such as water and chlorine were removed by passing the gas through the liquid-air trap *L*, in which the oxygen could also be condensed and stored.

Reactions were carried out as follows. Oxygen at a pressure of (usually) 100 mm. was left in contact with the catalyst at the reaction temperature for at least half an hour and usually longer. As is shown elsewhere this gives an adsorbed layer covering about 0.7 of the surface. The reaction vessel was pumped for exactly $1\frac{1}{2}$ min. in the case of experiments at 263° C or 1 min. in experiments at higher temperatures. The gas under examination was then admitted, and the course of the reaction with

time followed by the change in pressure on the manometer M_1 . The quantity of gas could be measured elsewhere in the apparatus (in the main vacuum line or the analyser A) before admission to the reaction vessel. After a certain time the contents of the reaction vessel were analysed by passing them into the U-tube B which was cooled in liquid air. Any non-condensable gas was measured on the manometer M_1 .

ANALYSIS

The analyser (figure 1) consists of a bulb A of 130 c.c. capacity to the bottom of which is attached a closed capillary tube D of 1.5 mm. bore. The tube D is covered with two thicknesses of asbestos paper on which is wound a coil of nichrome wire which can carry a small heating current; outside this is more asbestos paper. The gases to be analysed are first condensed by means of liquid air in the U-tube B and the analyser is pumped for a few minutes to remove any non-condensable gas. The gases are allowed to vaporize and then condensed in the tube D by cooling it in liquid air. When D is thoroughly cooled, the liquid air is removed, and D is surrounded with a thick wad of cotton-wool and allowed to warm up slowly; the pressure is measured on the manometer M_2 at suitable intervals of time and a step-wise graph of pressure against time is thus obtained. The cotton-wool is removed after most of the ethylene oxide has vaporized and the heating current is started 2 min. later in order to vaporize the water.

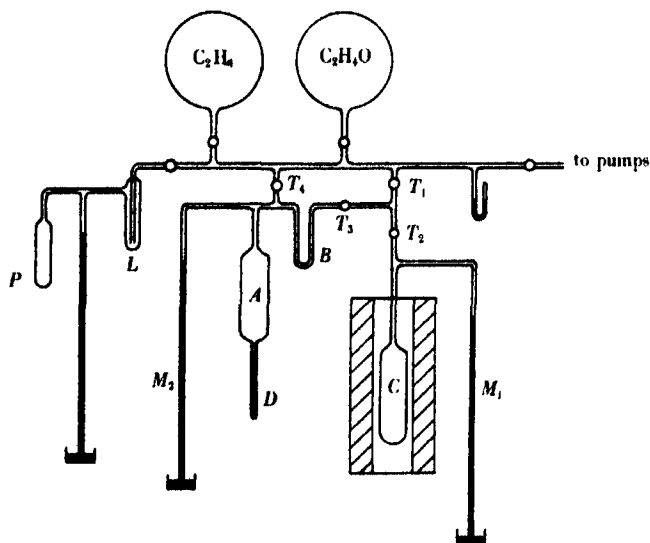


FIGURE 1. Apparatus.

This method of analysis has proved very convenient and accurate. A sample analysis with an artificially made up mixture of ethylene, carbon dioxide, ethylene oxide and water is shown in figure 2. The cotton-wool was removed at A , and the heating current was started at B . Separation of the components of the mixture was

always sharp except between carbon dioxide and ethylene oxide, where it was sometimes found that there was no completely horizontal portion of the curve. From trial mixtures it was found that the midpoint *P* of the flattened portion of the curve between the carbon dioxide and the ethylene oxide should be taken as their point of separation. The maximum error in the analysis of each component was ± 0.2 mm.; except in the case of water which tended to be low when there was much present.

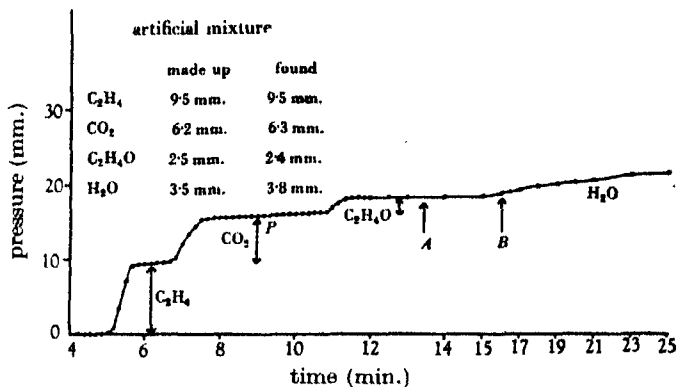


FIGURE 2. Sample analysis by micro-fractionation.

TABLE 1

compound	boiling point (° C) at			
	5 mm.	10 mm.	20 mm.	760 mm.
CO	-220	—	-213	-191
CH ₄	—	—	< -200	-161
C ₂ H ₄	-159	-153	-148	-104
C ₂ H ₂	-148	-142	-137	-88
C ₂ H ₆	-135	-130	-126	-84
CO ₂	-126	-121	-115	-78
H. CHO	—	-90	—	-21
C ₂ H ₄ O	-71	-63	-54	11
CH ₃ .CHO	-70	-61	-51	19
CH ₃ .OH	-25	-15	-5	65
H ₂ O	1	11	22	100
H. COOH	-5	3	—	101
CH ₃ .COOH	—	18	30	119

This method of analysis was capable of separating or detecting all the possible simple compounds that might be formed in the reaction between ethylene and oxygen. These are listed in table 1 together with their boiling points at 5, 10, 20 and 760 mm. The quantity of any one substance present was such that its pressure in the analyser was generally less than 15 mm. From the table it appears that substances should be separable when the boiling points at 5 mm. pressure differ by more than 10° C. Methane and carbon monoxide are non-condensable in liquid air; the

remaining substances should fall into the following groups: (a) ethylene and ethane, (b) acetylene and carbon dioxide, (c) formaldehyde, (d) ethylene oxide and acetaldehyde, (e) methyl alcohol, (f) water, formic and acetic acids. Ethane or acetylene though not accurately measurable would produce a lack of sharpness in the separation between ethylene and carbon dioxide; none was found in the reactions examined here. Ethylene oxide and acetaldehyde were indistinguishable. Water would not be separable from formic and acetic acids, but the presence of these acids in any quantity would be shown by a discrepancy in the total carbon analysis. The relationship between the volume of the analyser and that of the reaction vessel was determined at the various reaction temperatures by using ethylene after the catalyst had been deoxygenated; no ethylene is adsorbed on the catalyst under these conditions. In this way a complete analysis of the products of reaction could be made.

RESULTS

(a) *Ethylene*

The carbon dioxide which is produced in the reaction between gaseous oxygen and ethylene may arise in two ways: (i) from the further oxidation of ethylene oxide or (ii) directly from ethylene without passing through ethylene oxide. It has already been seen that part, at least, of the carbon dioxide originates through the first process; the reaction between ethylene and an adsorbed oxygen film enables one to decide whether the second process is operative also.

The direct reaction to produce carbon dioxide is



accompanied by an increase in gas pressure, while with the reaction to ethylene oxide,



no increase in pressure will result. Consequently, when ethylene is admitted an increase in pressure should take place at once if the direct reaction is present, whereas, if it is absent, there should be an induction period in the pressure-time curve. Figure 3 shows a graph of increase in pressure (δp) against time for the clean-up of an oxygen layer by ethylene; the complete absence of an induction period shows that direct oxidation of ethylene to carbon dioxide can take place, through intermediates of short life.

The products of the reactions between ethylene and adsorbed oxygen were analysed as described above. Sample analyses are shown in table 2; all values are in millimetres pressure in the reaction vessel. The second column in the table gives the initial pressure of ethylene in the reaction vessel, and the third column, that of oxygen in the one experiment where gaseous oxygen was used. Under the heading Δp is listed the increase in pressure during reaction. The penultimate column gives the total carbon analysis ($\Sigma \text{C}_2 = \text{Et} + \frac{1}{2}\text{CO}_2 + \text{EtO}$), while $\Delta \text{O}_2 = \text{CO}_2 + \frac{1}{2}\text{EtO} + \frac{1}{2}\text{H}_2\text{O}$ is the amount of oxygen removed from the catalyst.

The first point of interest is the smallness of the quantity of ethylene oxide formed compared with that of the carbon dioxide. To check that this was not due to an anomaly in the catalyst, a mixture of ethylene and gaseous oxygen was allowed to react on the catalyst for 11 min. (experiment 237); the results were in agreement with those of flow experiments. The first three experiments were taken to completion as judged by no further increase in pressure occurring. In experiments 312 and 338 only about half the oxygen was removed from the surface before analysis after 2 min. reaction. The quantity of ethylene oxide is larger than in those experiments in which the reaction was taken to completion, showing that part of the ethylene oxide formed is subsequently destroyed.

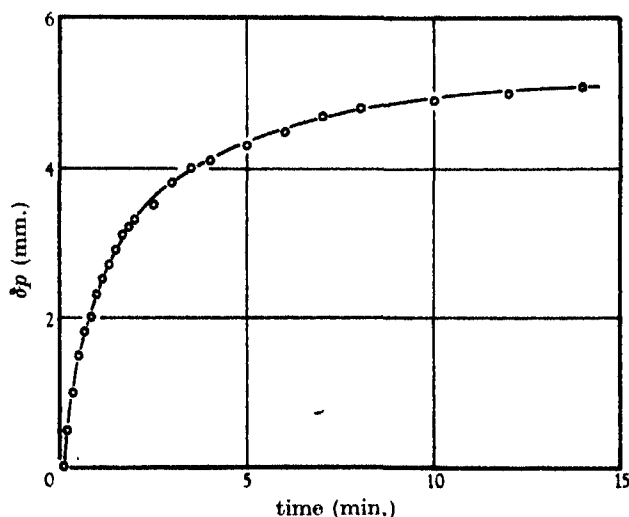


FIGURE 3. Pressure-time curve for reaction of ethylene with an adsorbed oxygen layer.

TABLE 2. 263° C

experi- ment number	p_{Et}	p_{O_2}	Δp	Et	CO_2	EtO	H_2O	ΣC_2	ΔO_2
233	31.1	—	4.9	27.7	5.1	0.4	2.8	30.7	6.7
234	24.0	—	4.9	21.5	4.9	0.7	2.4	24.7	6.5
235	14.1	—	5.1	11.2	4.7	0.6	2.7	14.15	6.4
237	27.0	69.3	-2.0	19.1	9.1	5.5	9.5	29.2	(16.6)
238	28.0	—	5.0	24.5	4.8	0.8	2.9	27.7	6.7
312	30.4	—	2.8	27.4	2.7	1.4	1.8	30.2	4.3
338	30.2	—	3.0	26.9	2.4	1.3	2.2	29.4	4.2

Another feature shown in table 2 is the discrepancy between the carbon and hydrogen analyses. According to the equation $C_2H_4 + 6O_{ads} \rightarrow 2CO_2 + 2H_2O$, the quantities of carbon dioxide and water formed should be equal, yet in all clean-up experiments the quantity of water was less than that of CO_2 by about 40 %. This feature was never found with ethylene-oxygen mixtures and is not an analytical

discrepancy; it appears due to adsorption of hydrogen on the catalyst. The following experiments were made to trace this discrepancy. The catalyst was oxygenated in the usual way—by leaving it in contact with 100 mm. of oxygen at 263° C for a sufficient time. After pumping out the reaction vessel for 1½ min., a measured quantity of ethylene was admitted and the clean-up allowed to go to completion. The gases were then removed for analysis, and, without further pumping of the reaction vessel, oxygen was put in contact with the catalyst and the temperature raised, being kept at 340° C for over an hour before cooling back to 263° C. The contents of the reaction vessel were pumped through the trap *B* (figure 1) immersed in liquid air, and the carbon dioxide and water were separated from the oxygen and analysed. Thus any deposit on the catalyst could be estimated. The results are shown in table 3 (all figures are in pressures of mm. of mercury). The last column gives the total hydrogen analysis, $\Sigma H_4 = Et + EtO + \frac{1}{2}H_2O$. The discrepancy between the carbon dioxide and the water is again seen in the clean-up, but practically all the hydrogen which is lost here is recovered on treatment with oxygen. The total analyses for carbon and hydrogen agree with the amount of ethylene put in to the degree of accuracy of measurement. It appears from experiments 549 and 551 that there is not much adsorption of carbon; when ethylene is put in contact with the catalyst after it has been cleaned with ethylene, it is not further adsorbed. It seems probable that this carbonaceous deposit arises from ethylene oxide which has been found to be strongly adsorbed and that ethylene itself is adsorbed only very slightly if at all. The hydrogen left on the catalyst after the clean-up does not appear to be adsorbed as water, since when water was admitted to either a cleaned or oxygenated catalyst none was adsorbed. From the relative quantities of carbon and hydrogen adsorbed (experiments 549 and 551) it seems unlikely that all the hydrogen is associated with the carbon on the catalyst, and it must therefore be present alone on the catalyst; as will be shown later, this catalyst adsorbs hydrogen strongly. It seems probable that the hydrogen adsorbed here comes from the breakdown of one of the intermediates between ethylene and carbon dioxide.

TABLE 3

experiment number	$p_0 Et$	Et	CO_2	EtO	H_2O	ΣC_2	ΣH_4
548 clean-up	56.3	52.1	5.8	0.5	3.4	55.5	54.3
549 oxygen treatment	—	—	1.2	—	3.8	0.6	1.9
	56.3	—	—	—	—	56.1	56.2
550 clean-up	22.3	19.1	5.0	0.3	2.5	21.9	20.65
551 oxygen treatment	—	—	0.6	—	3.4	0.3	1.7
	22.3	—	—	—	—	22.2	22.35

(b) Acetaldehyde

When acetaldehyde was admitted to an oxygenated catalyst, no change in pressure was found. Analysis, however, showed the presence of carbon dioxide and water. This must mean that the oxidation is extremely rapid and must have occurred within

the time taken to close the tap to the reaction vessel and read the manometer, i.e. in less than 5 sec. This is to be compared with about 5 min. taken to clean up 90 % of the oxygen with ethylene. This rapid reaction is in accordance with all the other information gained about acetaldehyde.

It was also found that acetaldehyde had no further action on a clean catalyst, except that it was adsorbed to form a deposit on the catalyst, which could be burned off in oxygen. The quantity of deposit was much greater than in the case of ethylene. After the oxygenated catalyst had been treated with acetaldehyde for 6 min. followed by another 6 min. in contact with fresh acetaldehyde, the deposit was burned off in oxygen, and 17.0 mm. of carbon dioxide and more than 7 mm. of water were recovered. This is to be compared with 1.2 mm. of carbon dioxide and 3.8 mm. of water burned off the catalyst after treatment with ethylene for 21 min.

The same discrepancy between the amounts of carbon dioxide and water was found as in the case of ethylene. In one experiment on the clean-up by acetaldehyde 9.5 mm. of carbon dioxide and 4.8 mm. of water were obtained; in another, the quantities were 7.4 and 4.6 mm. respectively. Again it would seem that hydrogen is being adsorbed by the catalyst.

A certain small amount of gas, non-condensable in liquid air, was always formed when acetaldehyde was in contact with the catalyst; this measured about 0.4 mm. after 6 min. It is probable that this is a mixture of carbon monoxide and methane from the decomposition of acetaldehyde, $\text{CH}_3\text{CHO} \rightarrow \text{CH}_4 + \text{CO}$. It is known that acetaldehyde decomposes thermally in this way at temperatures of the order of 500° C (Hinshelwood & Hutchison 1925).

(c) Ethylene oxide

Interesting results were obtained with ethylene oxide. When it was put in contact with an oxygen-covered catalyst, the total increase in pressure was much greater than would be expected from the quantity of oxygen on the catalyst. Furthermore, with a deoxygenated catalyst, reaction occurred with an increase in pressure. The cause of this was seen when analysis disclosed ethylene among the products of reaction (table 4). All figures are in mm. pressure of mercury. In the first of these experiments, ethylene oxide was admitted to a catalyst which had been covered with an adsorbed layer of oxygen in the usual manner; the reaction was allowed to proceed for 15 min., when, as judged from there being no further increase in pressure, it had reached completion. In the second experiment, fresh ethylene oxide was admitted to the cleaned catalyst and after 19 min. the reaction had reached completion.

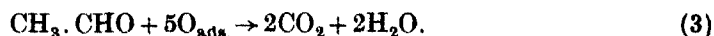
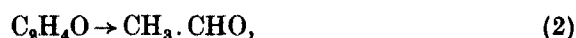
TABLE 4

experiment number	<i>p</i>	Δp	N.C.	Et	CO ₂	EtO (<i>A</i>)	H ₂ O
241	21.8	10.1	0.1	9.3	8.6	6.0	7.9
242	28.3	6.3	0.5	14.4	3.1	11.2	5.4
243	28.4	4.0	0.9	9.0	2.6	18.1	2.7

The third experiment was similar to the second except that, owing to the greatly reduced activity of the catalyst, the reaction had not reached completion after 20 min. when the analysis was made.

In all three cases ethylene is produced. This must come from the ethylene oxide by the reaction $C_2H_4O \rightarrow C_2H_4 + O_{ads}$.—a reversal of the reaction by which ethylene oxide is formed. In experiments 242 and 243, where there is no oxygen on the catalyst initially, this adsorbed oxygen is the only source of oxygen for the production of the carbon dioxide and water. The column headed EtO (A) gives the amount of ethylene oxide and acetaldehyde found at the end of the experiment—these are indistinguishable by the analytical technique used here. Since the first two reactions have been taken to completion, and since, as has been seen, acetaldehyde is without action on a clean catalyst, the substance listed under this heading in table 4 must be acetaldehyde alone. In the third experiment, where the reaction is not complete, it is a mixture of ethylene oxide and acetaldehyde. This furnishes another proof of the isomerization of ethylene oxide to acetaldehyde.

A fairly complete picture is now presented of what happens when ethylene oxide reacts with a deoxygenated catalyst. The following reactions occur on the surface of the catalyst:



Part of the ethylene oxide is decomposed into ethylene and adsorbed oxygen; part isomerizes to acetaldehyde which then reacts with the adsorbed oxygen to produce carbon dioxide and water. This cycle of operations proceeds until all the ethylene oxide is used up, and the products consist finally of ethylene, acetaldehyde, carbon dioxide and water. Since reaction (3) is known to be very fast, the quantity of oxygen present on the catalyst at any time during the experiment is very small. The relative rates of reactions (1) and (2) can be gauged from the fraction of the initial ethylene oxide which has been turned into ethylene. From experiment 242 above, and other experiments, this fraction is somewhat less than one-half. Hence it can be said that the velocity of reaction (1) is of the same order, though slightly slower than that of reaction 2.

When there is only a trace of oxygen adsorbed on the catalyst, as above, all the carbon dioxide will be produced via acetaldehyde. When, however, one starts with an oxygenated catalyst, there is the possibility of an additional reaction—a direct oxidation of ethylene oxide not via acetaldehyde:



It is impossible to separate these two reactions with certainty, but assuming that reactions 1 and 2 take place only on the clean silver surface, a qualitative answer can be obtained. If reaction 4 is absent, then the ratio of ethylene to acetaldehyde should be constant, irrespective of whether the surface was originally covered with

oxygen or not. That is, at the completion of reaction, the fraction of the original ethylene oxide which is converted into ethylene should be the same in both cases. If direct oxidation of ethylene oxide occurs, then the amount of ethylene should be less when the catalyst was originally covered with oxygen. In experiment 241 above (oxygen covered catalyst) the fraction of the original ethylene oxide converted to ethylene is 0.43; in experiment 242 (clean catalyst) it is 0.51. In a similar pair of experiments this fraction is 0.37 for an oxygenated catalyst, and 0.41 for a clean catalyst. Thus there is a suggestion that a small amount of ethylene oxide is oxidized directly to carbon dioxide and water, although the main route for this oxidation is via acetaldehyde. This direct oxidation can be significant only when the amount of oxygen on the catalyst is large, as its rate is proportional to the oxygen concentration on the surface, whereas that of the competing reaction is determined by the rate of isomerization, and this is, in all probability, independent of the oxygen concentration.

The pressure-time curves for the clean-up of the oxygenated catalyst by ethylene oxide have an unusual shape. In figure 4 the increase in pressure δp is graphed against time for the reaction of ethylene oxide (*a*) with an oxygenated catalyst, and (*b*) with a deoxygenated catalyst. With the deoxygenated catalyst, the rate of reaction falls steadily as the ethylene oxide is used up. When there is oxygen initially on the catalyst, however, the rate decreases steadily up to the point *P* where it increases; after the point of inflexion, the rate again decreases steadily. The most reasonable explanation of this phenomenon is that up to the point *P* there is enough oxygen on the surface of the catalyst to oxidize the acetaldehyde as fast as it is formed by isomerization; up to this point, the rate of reaction is controlled by the rate of isomerization, and there is very little free acetaldehyde in the gas phase. At *P* both the surface oxygen concentration and the acetaldehyde concentration are low, and the rate of reaction is small. After the point *P*, the acetaldehyde concentration begins to build up, and this increase more than compensates for the low adsorbed oxygen concentration, so that after *P* there is an increase in rate of reaction. The rate of reaction is now controlled by the rate at which ethylene oxide is broken down into ethylene and adsorbed oxygen (reaction 1), and decreases with time as in (*b*) owing to the ethylene oxide being used up. Exact kinetic measurements to establish this hypothesis are not possible, as ethylene oxide and acetaldehyde are adsorbed on the catalyst to form an organic deposit which reduces the activity of the catalyst.

In one experiment (537) ethylene oxide was admitted to the oxygenated catalyst for 1 min. which was the time required to reach the point of inflexion. The quantity of ethylene found by analysis showed that the first part of the reaction, up to the point of inflexion, is not due to a rapid reaction between ethylene oxide and adsorbed oxygen. In this experiment, 2.6 mm. of ethylene oxide were transformed into ethylene, and 4.1 mm. to carbon dioxide. In the following experiment, in which fresh ethylene oxide was allowed to react on the now largely deoxygenated catalyst, at completion 18.0 mm. had gone to ethylene and 18.9 mm. to acetaldehyde or via it to carbon dioxide. Thus in experiment 537, if the carbon dioxide is all produced via acet-

aldehyde, one would have expected about 4.1 mm. of ethylene. When allowance is made for the reoxidation of part of this ethylene, and for the fact that the reaction



is probably slower on an oxygen covered surface than on a clean surface, it is thus very probable that almost all the carbon dioxide formed in experiment 537 has come via acetaldehyde.

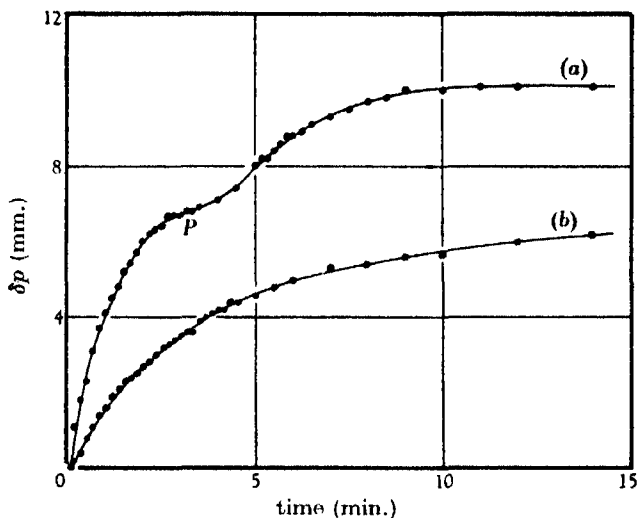


FIGURE 4. Pressure-time curve for reaction of ethylene oxide with (a) an adsorbed oxygen layer, (b) a clean catalyst.

A number of experiments were made to check the analyses and also to find out the quantity of organic deposit on the catalyst. Complete analysis showed the same discrepancy between the water and the carbon dioxide as was found in the case of ethylene. When the freshly oxygenated catalyst is first treated with ethylene oxide, the amount of water in the products is less than equivalent to the carbon dioxide; with further treatments the reverse is true. The discrepancy is due to adsorption by the catalyst, and the nature of the adsorbed layer changes as it increases in amount. The first deposit consists largely of hydrogen, probably in an elementary form, which later on is displaced by a more carbonaceous layer. The deposit has thus no very definite composition; in quantity it is very much greater than that formed from ethylene.

At the end of each series of experiments, the deposit was burned off by heating the catalyst in oxygen at a temperature of 270–300° C overnight. Measurement of the carbon dioxide and water formed, together with the analyses of the products of each experiment, enabled total carbon and hydrogen analyses to be made. The very good agreement which was found between these and also between these and the

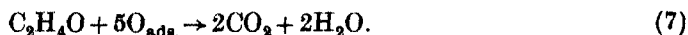
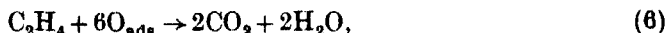
amounts of ethylene oxide admitted to the reaction vessel justifies one in thinking that the substances analysed have been correctly identified.

The effect of pretreating the catalyst with hydrogen was examined. In this treatment all the adsorbed oxygen was removed, and in addition 9.1 mm. of hydrogen were adsorbed. This adsorbed hydrogen had little effect on the reactions occurring, except that it combined with some of the oxygen formed from the ethylene oxide, with the result that the amount of carbon dioxide produced was less than in other experiments. A total of 8.1 mm. of the adsorbed hydrogen was recovered.

KINETICS OF THE REACTION BETWEEN ETHYLENE AND ADSORBED OXYGEN

It is possible, in an approximate way, to use the increase of pressure with time, which occurs when ethylene reacts with an adsorbed layer of oxygen, to examine the kinetics of the direct reaction of ethylene to carbon dioxide.

The three main reactions can be expressed by the overall equations



Of these, the second and third produce an increase in pressure. Since the amount of ethylene oxide formed has been shown by analysis (table 2) to be small, the increase in pressure is a measure mainly of the second reaction.

Equation (6) expresses only the beginning and end of this reaction, and it must be made up of a series of steps. Since none of the intermediates is produced in detectable amounts, all of these steps except the first are relatively fast, and the rate of production of carbon dioxide is the rate of this first slow step, which can be represented by the equation



It is the kinetics of this step which will be obtained from measurements on the increase in pressure.

The first relationship sought is that between the rate of reaction and the fraction θ of the surface covered by oxygen. The latter is related to the increase in pressure, but owing to the fact that some of the hydrogen in the ethylene is retained by the catalyst as noted previously, this relationship is complex. It can be derived if the not unreasonable assumption be made that a constant fraction α of the hydrogen from the ethylene is returned to the gas phase as water. Then, if P_1 and p_1 represent the pressures of ethylene initially and at time t respectively, p_2 the pressure of ethylene oxide, and p_3 ($= \frac{1}{2}\text{CO}_2$, since only a negligible quantity of carbon is adsorbed) the quantity of ethylene transformed into carbon dioxide and water, the following equations hold:

$$P_1 = p_1 + p_2 + p_3, \quad p = p_1 + p_2 + 2p_3 + 2\alpha p_3,$$

where p is the total pressure at time t . Hence the increase in pressure is

$$\delta p = p - P_1 = (1 + 2\alpha) p_3,$$

and at completion

$$\Delta p = (1 + 2\alpha) p_{\infty}.$$

If R is the quantity of oxygen measured in mm. pressure in the reaction vessel required completely to cover the catalyst surface, the quantity of oxygen initially on the catalyst is

$$\begin{aligned} R\theta_0 &= R\theta + \frac{1}{2}p_2 + 2p_3 + \alpha p_3 \\ &= R\theta + \frac{1}{2}p_2 + \left(\frac{2+\alpha}{1+2\alpha}\right)\delta p \\ &= \frac{1}{2}p_{\infty} + \left(\frac{2+\alpha}{1+2\alpha}\right)\Delta p. \end{aligned}$$

Hence

$$R\theta = \left(\frac{2+\alpha}{1+2\alpha}\right)(\Delta p - \delta p) + \frac{1}{2}(p_{\infty} - p_2).$$

Since the quantity of ethylene oxide formed is always small, this can be written approximately

$$R\theta = \left(\frac{2+\alpha}{1+2\alpha}\right)(\Delta p - \delta p) = g(\Delta p - \delta p). \quad (8)$$

The value of g is obtained from the equation

$$R\theta_0 = g\Delta p, \quad (9)$$

where $R\theta_0$ is determined from analysis of the products of reaction at the end of the experiment. From table 2 it can be seen that slightly more than half the hydrogen appears in the gas phase as water. Taking α as 0.6, g has the value 1.18 which is in rough agreement with the values of g found in the experiments of table 6.

These equations can now be applied to reaction 6. The quantity of carbon dioxide formed via ethylene oxide is small enough to be neglected, and so the kinetic equation for reaction 6 can be written approximately

$$\frac{dp_3}{dt} = B_1 f_1(p_1) f_2(\theta). \quad (10)$$

It will be shown that this equation is actually

$$\frac{dp_3}{dt} = B_1 p_1 \theta^2. \quad (11)$$

Substituting from above, this becomes

$$\frac{1}{1+2\alpha} \frac{d}{dt}(\delta p) = B_1 \frac{g^2}{R^2} p_1 (\Delta p - \delta p)^2$$

or

$$\frac{d}{dt}(\Delta p - \delta p) = -K p_1 (\Delta p - \delta p)^2, \quad (12)$$

where $K = (1 + 2\alpha) B_1 \frac{g^2}{R^2}$. On integration this becomes

$$\frac{1}{\Delta p - \delta p} - \frac{1}{\Delta p} = K \int p_1 dt \doteq K p_1 t, \quad (13)$$

since the change in the ethylene pressure during an experiment is relatively small. Consequently, if the kinetic equation (10) has the form assumed (11), a straight line should be obtained on plotting $1/(\Delta p - \delta p)$ against time.

Figure 5 shows a typical result; straight lines were obtained over a wide range. There was a general tendency for the line to curve upwards towards the end of the experiments when θ had fallen to below about one-fifth of its original value; this is not surprising in view of the approximations made and may indicate that the oxidation via ethylene oxide is more important when θ becomes small. The experiments thus show that the rate of the direct oxidation of ethylene to carbon dioxide is proportional to the square of the adsorbed oxygen concentration.

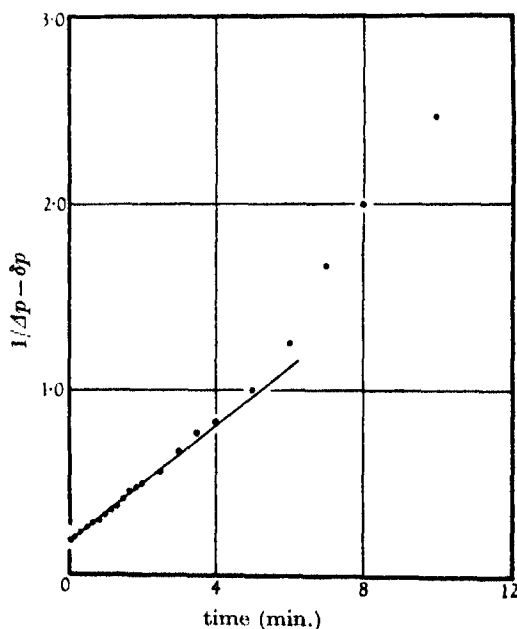


FIGURE 5. Kinetics of reaction of ethylene with an adsorbed oxygen layer.

Experiments made at varying ethylene pressures established that approximately the rate of reaction was proportional to the ethylene pressure.

A number of experiments between ethylene and adsorbed oxygen films was made in which the quantity of oxygen on the catalyst initially was varied. This was done by allowing the catalyst to come to equilibrium with varying pressures of oxygen, ranging from 3 to 750 mm. (table 5). The unexpected result was found, that the

velocity constant K of equation (13), though constant in any one experiment, was dependent on the initial value of θ ; the value of K falls as θ_0 increases. This is illustrated by the figures in table 5, particularly those at temperatures of 285 and 302° C. This phenomenon appears to be a characteristic of the silver catalyst, as it has also been found in experiments on the rate of evaporation of oxygen. It is illustrated again by one experiment (384) in which an oxygen film ($R\theta_0 = 7.5$ mm.) was evaporated into vacuum until the adsorbed concentration had fallen to $R\theta = 3.7$ mm.; the film was then cleaned up with ethylene at 263° C, and the velocity constant K for this reaction was 4.34×10^{-3} corresponding not to $R\theta = 3.7$ but to $R\theta$ between 7 and 9 judged from table 5. This type of behaviour has been found in other kinetic experiments with adsorbed oxygen.

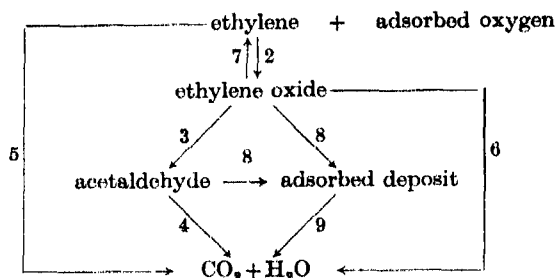
TABLE 5

experiment number	p_1 (mm.)	$R\theta$ (mm.)	Δp (mm.)	g	$K \times 10^3$	$B_1 \times 10^3$
1. $T = 263^\circ \text{C}$						
385	16.0	3.65	3.0	1.22	9.0	20.2
377	25.0	5.3	4.0	1.32	6.4	11.0
375	27.6	6.05	5.1	1.19	10.5	24.8
373	21.5	7.0	5.8	1.21	7.2	16.4
379	23.0	7.3	6.4	1.14	8.5	21.8
382	19.9	9.65	8.4	1.15	3.6	9.1
383	18.6	10.25	8.7	1.18	3.23	7.7
2. $T = 285^\circ \text{C}$						
390	35.0	4.0	3.0	1.33	16.6	31.4
391	25.9	4.2	3.1	1.35	18.9	34.6
397	13.4	4.9	4.2	1.17	16.6	40.5
395	10.6	6.5	5.5	1.18	12.3	29.4
392	31.4	6.8	5.5	1.24	11.6	25.2
394	35.2	9.1	8.1	1.12	7.4	19.5
3. $T = 302^\circ \text{C}$						
400	12.2	3.0	2.3	1.30	43.5	86
403	26.4	3.25	2.9	1.12	23.8	63
402	23.9	5.9	5.0	1.18	15.5	37.2
405	33.8	6.0	4.7	1.27	16.3	33.7
404	53.5	8.5	7.2	1.18	8.8	21.1

An approximate value was found for the energy of activation of the reaction between ethylene and adsorbed oxygen. The constant B_1 was derived from K by the formula given above; R was determined from measurements of the adsorption isotherms of oxygen; it equals 10.0 mm. at 263° C. (This work will be published shortly.) From the values of B_1 corresponding to similar values of $R\theta_0$ at the three temperatures, estimates of the energy of activation were made. These can only be very approximate, but the value of 10 ± 3 kcal. can be accepted as substantially correct.

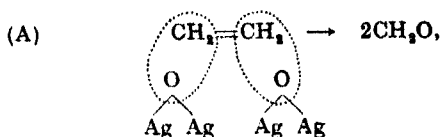
DISCUSSION

These experiments with oxygen films have established the following scheme of reactions:

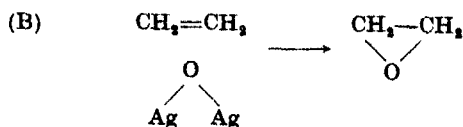


When ethylene reacts with oxygen adsorbed on the catalyst, part of it is converted by the addition of a single oxygen atom into ethylene oxide (2); another part of it is oxidized to carbon dioxide and water (5), via some intermediates which have not been isolated but which are presumably extremely easily oxidized. The rate of this direct reaction is proportional to the square of the concentration of adsorbed oxygen atoms. There is no way of finding from the present experiments the kinetic order of the reaction producing ethylene oxide, but it seems reasonable to assume that the reaction is first order with respect to the adsorbed oxygen concentration. This difference in kinetics accounts for the fact that the ratio of ethylene oxide to carbon dioxide in the clean-up experiments is less than that found in reactions between ethylene and gaseous oxygen (cf. Part III) when a considerable amount of oxygen has been removed from the catalyst. In the clean-up experiments during most of the reaction there is more oxygen on the catalyst than during reactions between ethylene and oxygen, thus favouring the direct reaction to carbon dioxide.

The kinetics of the direct oxidation (5) suggest that the reaction involves two adsorbed oxygen atoms per ethylene molecule. It may be possible to represent this process as follows:



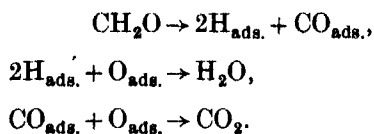
whereby two formaldehyde molecules are formed, which are then further oxidized. When ethylene meets a single oxygen atom, ethylene oxide is formed, viz.



It is probable that the ethylene in these reactions is not adsorbed on the catalyst, or only very weakly adsorbed. There is no measurable adsorption when ethylene

is put in contact with the catalyst, nor is there any appreciable carbonaceous deposit recoverable on heating the catalyst in oxygen. Thus the reactions represented by (A) and (B) are between gaseous ethylene, or at most ethylene adsorbed by weak van der Waals forces, and adsorbed oxygen atoms. The fact that no hydrogenation of ethylene occurs on the silver catalyst, though it adsorbs hydrogen, supports this view. It has been shown (Twigg & Rideal 1940) that in hydrogenation, the ethylene requires to be adsorbed on the catalyst at two points by the opening of the double bond.

It is probable that the small amount of carbonaceous deposit found after the clean-up by ethylene comes from the ethylene oxide formed. The hydrogen taken up by the catalyst in this reaction must come from one of the intermediates in the direct oxidation (5), for if it came from the ethylene oxide it would be accompanied by more carbon. It therefore seems possible that the catalyst can remove hydrogen from one of the intermediates, and it is suggested tentatively that the series of reactions following on (A) above may be



The reactions of ethylene oxide are more difficult to unravel. The most surprising of these is the production of ethylene from ethylene oxide. This presumably takes place through a reversal of the reaction (B) by which it is formed.

The present experiments and those of Part I show that ethylene oxide is catalytically isomerized to acetaldehyde (3) which is capable of very rapid oxidation to carbon dioxide and water (4). The intermediate steps in this last reaction are not known, though it may be that the power of the catalyst to split off hydrogen plays an important part.

The question of whether there is any direct oxidation of ethylene oxide to carbon dioxide and water (6) is somewhat difficult to settle. Exact kinetic measurements are not possible owing to the complexity of the reactions and also to the fact that the catalyst is continuously being covered with an organic deposit. It seems, however, that an approximate answer can be obtained by linking up the production of ethylene by reaction 7 with the isomerization (reaction 3). If it is assumed in the first place that these two reactions are not affected by the amount of oxygen on the catalyst, then an increase in the ratio of carbon dioxide plus acetaldehyde to ethylene in the case of an oxygenated catalyst, compared with this ratio for a deoxygenated catalyst, would indicate direct oxidation of the ethylene oxide. The increase found is small, and when allowance is made for a possible reduction in rate of reaction 7 when the catalyst is covered with oxygen, and also the reoxidation of the ethylene, then it appears that very little, if any, ethylene oxide undergoes direct oxidation to carbon dioxide; this is in agreement with the flow experiments (Part I). Thus, the existence of reaction 6 in the scheme outlined above is somewhat doubtful.

Ethylene oxide and also acetaldehyde give rise to an organic deposit on the catalyst. The composition of this appears to be ill-defined and also variable. When the catalyst is continuously treated with ethylene oxide, the ratio of hydrogen to carbon adsorbed falls; the hydrogen appears to be displaced by the carbon. This deposit can cover a considerable area of the catalyst; that recovered in one experiment would, on the basis of one carbon atom per silver atom, cover over 80 % of the total surface. This is reflected in the great reduction in rate in successive reactions of ethylene oxide on the catalyst.

The formation of this deposit provides an additional way by which carbon dioxide is produced. This route, however, is not of great importance. In a series of experiments (537-540) after about 160 min. total reaction time the equivalent of 17.5 mm. of carbon dioxide had formed; this is to be compared with the 8.2 mm. of carbon dioxide produced via other routes in 1 min. in experiment 537.

Thus the most important reactions which occur between adsorbed oxygen and the organic gases, ethylene, ethylene oxide and acetaldehyde are (a) the oxidation of ethylene to ethylene oxide and the reverse step, (b) the direct oxidation of ethylene to carbon dioxide and water, (c) the isomerization of ethylene oxide to acetaldehyde, and (d) the oxidation of acetaldehyde to carbon dioxide and water. These reactions occur in the absence of oxygen in the gas phase. Whether these are equally important when gaseous oxygen is present can only be shown by comparison with experiments in which ethylene and oxygen are both present in the gas phase.

REFERENCES

- Hinshelwood & Hutchison 1925 *Proc. Roy. Soc. A*, **111**, 380.
Steacie & Johnson 1926 *Proc. Roy. Soc. A*, **112**, 542.
Twigg & Rideal 1940 *Trans. Faraday Soc.* **36**, 533.

The mechanism of the catalytic oxidation of ethylene.

III. The reactions between ethylene and oxygen in a static system

By G. H. TWIGG, PH.D. *Department of Colloid Science, Cambridge*

(Communicated by E. K. Rideal, F.R.S.—Received 31 July 1945)

Static experiments have been carried out between ethylene and oxygen on the silver catalyst at different pressures and temperatures, and the rates of the several reactions occurring were determined. It was found at 263° C that (a) the rate of oxidation of ethylene to ethylene oxide is proportional to the ethylene pressure and to the first power of the concentration (θ) of adsorbed oxygen on the catalyst; (b) the rate of the direct oxidation of ethylene to carbon dioxide is proportional to the ethylene pressure and to the square of θ ; (c) the rate of oxidation of ethylene oxide is proportional to the ethylene oxide pressure and independent of θ .

The kinetic equations for the course of reaction with time, although incapable of analytical solution, were integrated by means of the Bush differential analyser, the numerical constants being fitted to one series of experiments. The goodness of fit and the accuracy of the predictions for other conditions were confirmation of the kinetics deduced.

These results confirm previous findings that the main steps in the reactions are (1) chemisorption of oxygen as atoms on the catalyst surface, (2) reaction of gaseous or weakly adsorbed ethylene either with one oxygen atom to form ethylene oxide, or with two atoms to form products which are oxidized rapidly to carbon dioxide, and (3) isomerization of ethylene oxide to acetaldehyde followed by the rapid oxidation of this to carbon dioxide. Energies of activation have been measured for these steps.

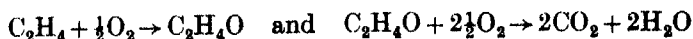
An attempt was made to measure directly the concentration of oxygen adsorbed on the catalyst during reaction, by cleaning it up with hydrogen after pumping out the reaction mixture; a value of 0.31 was found which is in fair agreement with that expected from the machine solutions.

In the previous paper, certain hypotheses were put forward regarding the reactions which occur when ethylene and oxygen interact in the presence of a silver catalyst. These hypotheses were based mainly on reactions between ethylene oxide, etc., and films of oxygen adsorbed on the catalyst. In the present experiments mixtures of ethylene and oxygen, and of ethylene oxide and oxygen, were allowed to react for a certain time and the products removed for analysis; in this way, kinetic equations were obtained which could be compared with those deduced from the reaction mechanisms postulated.

The apparatus and catalyst used were the same as were described in the previous paper. Before each series, i.e. experiments under the same conditions but for different reaction times, the catalyst was heated overnight in oxygen at the operating temperature; the gaseous oxygen was pumped out and part of the adsorbed oxygen removed from the catalyst by treatment with ethylene for 2 min. The ethylene-oxygen mixture was admitted and reaction allowed to proceed for some time before the reaction vessel was pumped out. In this way the quantity of oxygen adsorbed was regulated approximately to that obtaining during reaction. The series of experi-

ments was then started, those at the longest time being done first. Between each experiment, the reaction vessel was empty for about half an hour, but this appeared not to alter materially the surface oxygen concentration as the rate of evaporation is small. Failing this treatment, discordant results were obtained.

Theoretically, the reactions



could be followed by observing the change of pressure in the reaction vessel, the direct oxidation of ethylene to carbon dioxide and water involving no change of pressure. However, this was found unreliable owing to the large amount of oxygen adsorbed by the catalyst, and to the fact that some of the products of reaction are also adsorbed on the catalyst.

Analysis of the reaction products was carried out as described in the previous paper. The contents of the reaction vessel were passed through the U-tube (figure 1, previous paper) cooled in liquid air, where the residual oxygen was separated off. The quantity of this oxygen was measured by admitting the mixture to the analyser in successive portions, pumping the analyser out in between each admission. The quantity of oxygen determined in this manner agreed to within 2 % with that found indirectly by analysis of the other products.

During the separation from the oxygen, some of the condensable gas was lost by being swept through the U-tube before it had time to condense. Experiments with artificial mixtures showed that 3-9 % of the ethylene, ethylene oxide and carbon dioxide could be lost in this way. The amount lost depended on the proportion of oxygen present, but approximately the same fraction of each gas was lost in any one experiment; water was not lost in this way. A correction was made to the analysis for this loss; the discrepancy between ΣC_2 ($= \text{Et} + \text{EtO} + \frac{1}{2}\text{CO}_2$) and the quantity of ethylene put into the reaction vessel initially was divided up between the three gases in proportion to the amounts present, and added to the values found by analysis.

A further correction was made to the analyses on account of the unreacted gases in the connecting tubing of the reaction vessel, which were calculated to amount to 5 % of the total. A quantity equal to 5 % of the initial ethylene present was subtracted from the ethylene analysed. The results were finally expressed as a percentage of the original ethylene left as ethylene or transformed into ethylene oxide, carbon dioxide, or water. An example may make these corrections clear; the analyses are expressed as mm. pressure in the reaction vessel. The quantity of ethylene present at the beginning of the experiment was 20.6 mm.:

	Et	EtO	CO ₂	H ₂ O	ΣC_2
as analysed	9.9	4.8	9.1	9.6	19.25
corrected for loss in pumping	10.6	5.15	9.7	9.6	20.6
corrected for dead space	9.6	5.15	9.7	9.6	19.6
as a percentage	48.9	26.3	24.7	24.5	100

Analogous corrections were made in the case of reactions between ethylene oxide and oxygen.

RESULTS

Reactions between ethylene and oxygen

The series of reactions S1-S7 (table 1) were carried out at a temperature of 263°C; the initial pressures of the ethylene and oxygen were varied in order to determine the orders of the various reactions occurring. These experimental points are shown in figures 1-7; the full lines are obtained theoretically as explained later.

The deduction of the kinetics of the various reactions from these results is somewhat complicated owing to the fact that there are simultaneous and consecutive reactions occurring together. It is assumed that as suggested in the previous paper one component in each reaction exists in the gas phase, or is weakly adsorbed, and the other component is the catalyst surface or the oxygen film on it. It is also assumed that only the following four reactions need be taken into account:

(a) *Adsorption of oxygen*: $O_2 \rightarrow 2O_{ads}$. The rate of this step should be proportional to the oxygen pressure and to the square of the free catalyst surface. If p_0 is the oxygen pressure and θ the fraction of surface covered with oxygen, then the rate of this step is $A_0 p_0 (1 - \theta)^2$, less a term of the form $A'_0 \theta^2$ due to evaporation, which will be relatively small when ethylene is present.

(b) *Ethylene oxide production*: $C_2H_4 + O_{ads} \rightarrow C_2H_4O$. The rate of this step will be proportional to the ethylene pressure p_1 , and to some function $f_1(\theta)$ of the surface oxygen concentration, i.e. rate = $A_1 p_1 f_1(\theta)$.

(c) *Direct CO₂ production*: $C_2H_4 + 6O_{ads} \rightarrow 2CO_2 + 2H_2O$. The rate of this reaction is taken to be $B_1 p_1 \phi(\theta)$.

(d) *CO₂ via ethylene oxide*: $C_2H_4O \rightarrow CH_3 \cdot CHO \rightarrow 2CO_2 + 2H_2O$. Here the rate will be proportional to the ethylene oxide pressure p_2 and to another function of the adsorbed oxygen concentration, and can be represented as $A_2 p_2 f_2(\theta)$.

When these four reactions are combined, the kinetic equations are as follows:

$$-\frac{dp_1}{dt} = A_1 p_1 f_1(\theta) + B_1 p_1 \phi(\theta), \quad (1)$$

$$-\frac{dp_2}{dt} = A_2 p_2 f_2(\theta) - A_1 p_1 f_1(\theta), \quad (2)$$

$$-\frac{dp_0}{dt} \doteq A_0 p_0 (1 - \theta)^2. \quad (3)$$

These equations cannot be solved analytically, except in the special case when $f_1 = f_2 = \phi$. In order to apply them to the results, the assumption was made that θ was constant throughout the course of the reaction. This is not true, but it is sufficiently accurate to enable the quantities $A_1 f_1(\theta)$, etc., to be evaluated approximately, especially at the beginning of the reaction.

TABLE 1. Et + O₂. T 263° C

experiment number	time (min.)	% Et	% EtO	% $\frac{1}{2}$ CO ₂	% $\frac{1}{2}$ H ₂ O	% O ₂
S 1. Et = 20 mm.; O ₂ = 60 mm.						
313	45	21.9	18.2	59.9	—	36.9
314	25	36.2	27.5	36.0	33.9	59.3
315	16	48.9	26.0	24.7	24.5	70.8
316	8	70.2	18.7	11.1	11.4	85.7
317	4	80.1	14.2	5.7	5.7	91.9
318	2	90.1	7.1	1.8	2.8	96.5
S 2. Et = 20 mm.; O ₂ = 120 mm.						
319	35	9.7	20.4	69.8	—	63.3
320	20	27.5	30.1	42.3	39.8	76.6
321	12	44.1	29.8	25.9	24.9	84.6
322	7	59.1	26.0	14.9	15.1	90.3
323	4	74.9	16.5	8.5	7.4	94.6
324	2	86.2	9.2	4.6	4.5	97.0
S 3. Et = 20 mm.; O ₂ = 30 mm.						
325	55	47.7	9.7	42.6	—	11.9
326	30	56.1	15.7	28.2	25.6	40.3
327	18	64.6	17.2	18.2	16.6	59.1
328	10	75.9	14.9	9.2	9.1	76.6
329	6	85.2	9.3	5.5	5.8	85.8
330	3	91.1	7.3	1.6	2.5	94.3
S 4. Et = 40 mm.; O ₂ = 60 mm.						
332	55	42.1	15.5	42.4	—	10.5
333	30	52.0	21.0	27.0	25.9	40.2
334	18	64.0	19.9	16.1	15.3	61.6
335	10	75.4	15.8	8.8	8.9	77.2
336	6	83.1	11.3	5.6	5.3	85.5
337	3	90.3	7.3	2.4	2.6	92.5
S 5. Et = 10 mm.; O ₂ = 60 mm.						
339	35	6.1	21.3	72.6	—	61.6
340	20	24.4	32.5	43.1	42.1	75.7
341	12	42.1	29.7	28.2	27.7	83.5
342	7	58.5	25.6	15.9	15.9	89.9
343	4	73.2	18.0	8.8	8.8	94.1
344	2	86.2	10.2	3.6	2.1	97.6
S 6. Et = 40 mm.; O ₂ = 120 mm.						
345	45	17.3	24.8	57.8	—	37.9
346	25	31.6	32.6	35.8	—	58.7
347	16	44.6	31.6	23.8	22.4	71.3
348	8	64.4	23.9	11.7	11.5	84.4
349	4	80.0	14.2	5.8	5.8	91.8
350	2	89.5	7.5	3.0	3.3	95.7
S 7. Et = 20 mm.; O ₂ = 60 mm.						
351	45	19.7	22.8	57.5	—	38.9
352	25	32.9	28.3	38.8	37.2	57.1
353	16	47.1	28.4	24.5	23.8	70.8
354	8	66.8	21.3	11.9	11.9	84.4
355	4	80.6	14.5	4.9	5.4	92.4
356	2	90.0	7.9	3.9	2.1	95.3

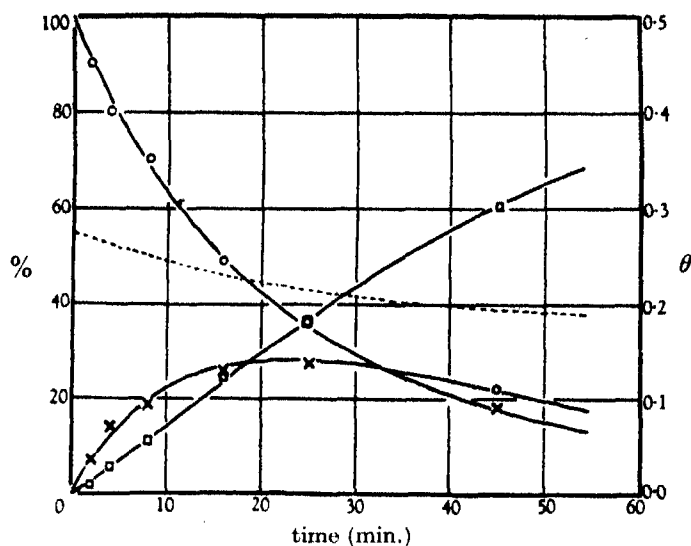


FIGURE 1

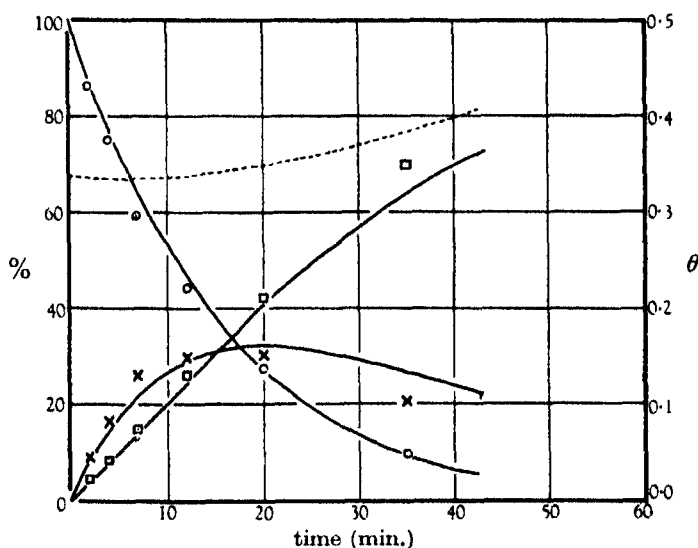


FIGURE 2

FIGURES 1-7. Course of reaction between ethylene and oxygen at 263° C.

Experimental points: O percentage ethylene remaining, x percentage ethylene converted to ethylene oxide, □ percentage ethylene converted to CO₂ and water.

Full lines: calculated course of reaction.

Dotted line: calculated surface concentration of oxygen (θ).

figure	series	p_{Et} (mm.)	p_{O_2} (mm.)
1	S1	20	60
2	S2	20	120
3	S3	20	30
4	S4	40	60
5	S5	10	60
6	S6	40	120
7	S7	20	60

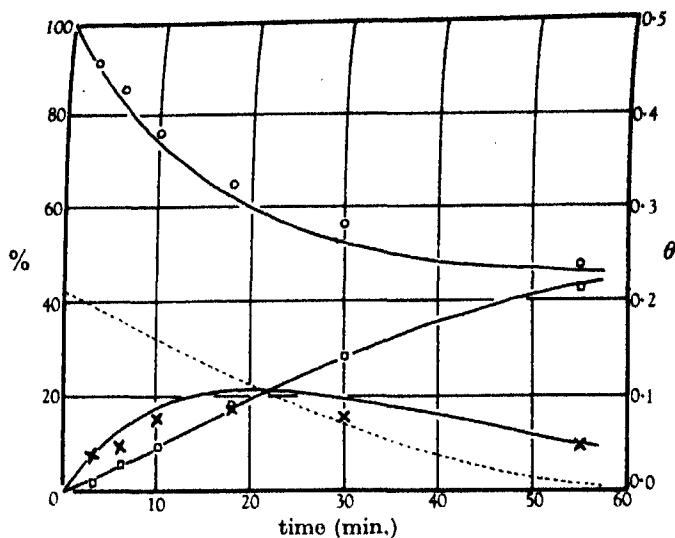


FIGURE 3

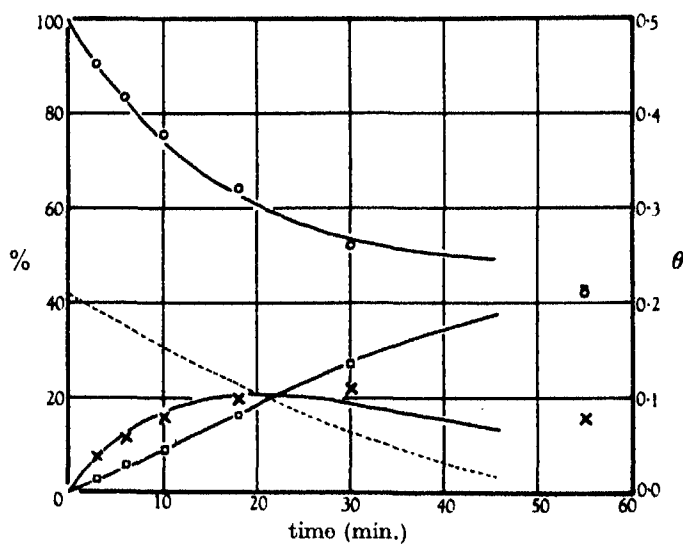


FIGURE 4

With this assumption, the solutions of the above equations are

$$1 + mn \frac{p_2}{p_1} = \left(\frac{P_1}{p_1} \right)^n, \quad (4)$$

$$\log_e \frac{P_1}{p_1} = (A_1 f_1(\theta) + B_1 \phi(\theta)) t, \quad (5)$$

$$\log_e \frac{P_0}{p_0} = A_0 (1 - \theta)^n t, \quad (6)$$

where

$$n = 1 - \frac{A_2 f_2(\theta)}{A_1 f_1(\theta) + B_1 \phi(\theta)}, \quad (7)$$

$$m = \frac{A_1 f_1(\theta) + B_1 \phi(\theta)}{A_1 f_1(\theta)}, \quad (8)$$

and P_1 and P_0 are the pressures of ethylene and oxygen at zero time. If $f_1 = f_2 = \phi$, then equation (4) is an exact solution independent of the behaviour of θ .

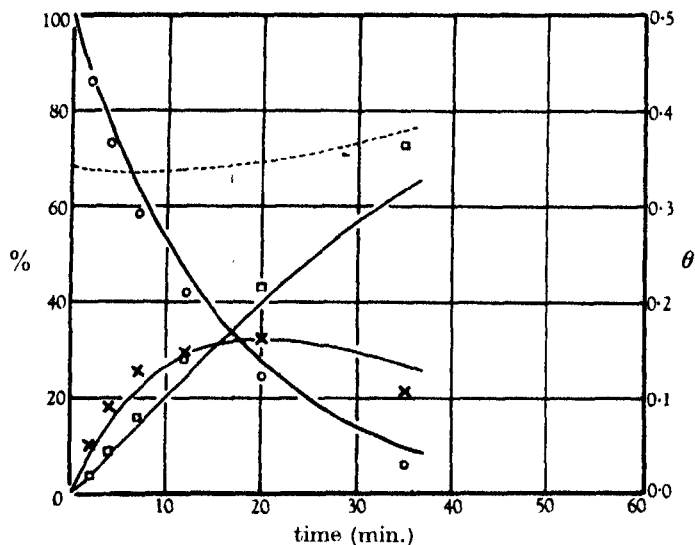


FIGURE 5

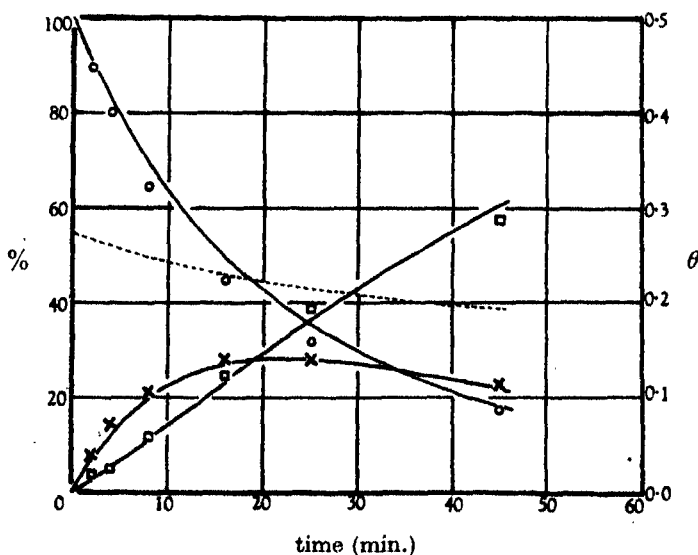


FIGURE 6

The quantities $A_1 f_1(\theta)$, etc., are evaluated as follows. Graphs of $\log p_0/P_0$ and $\log p_1/P_1$ against time are drawn and the gradients at the beginning of the reaction give $A_0(1-\theta)^2$ and $A_1 f_1(\theta) + B_1 \phi(\theta)$ respectively. Figure 8 shows the graph of

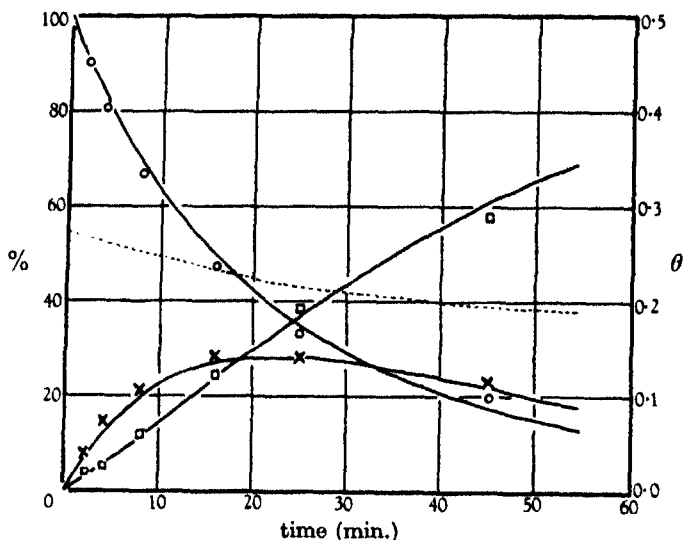


FIGURE 7

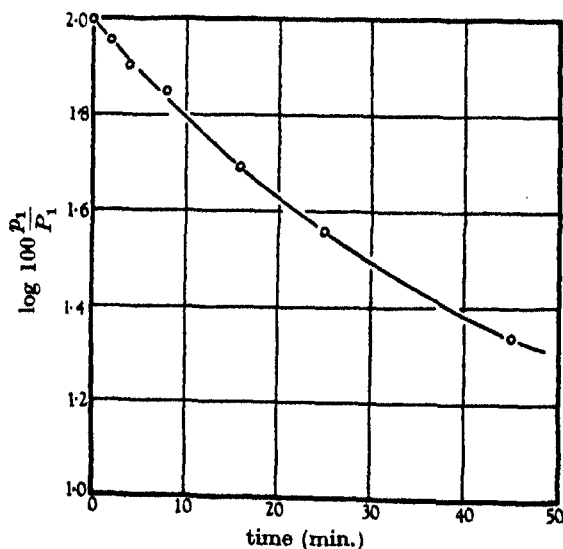


FIGURE 8. First order equation for disappearance of ethylene—series S1.

$\log p_1/P_1$ for the series S1; the upward curve of the line shows that θ is decreasing as the reaction proceeds. The quantities obtained in this way refer to the values of θ existing at the beginning of the reaction.

The separate values of $A_1 f_1(\theta)$ and $B_1 \phi_1(\theta)$ are found by estimating m . Equation (4) can be written

$$n \log \frac{P_1}{p_1} = \log \left(1 + mn \frac{p_2}{p_1} \right) = mn \frac{p_2}{p_1} - \frac{m^2 n^2}{2} \left(\frac{p_2}{p_1} \right)^2 + \dots$$

The higher terms of this expression can be neglected when this is applied to the early stages of reaction, since p_2/p_1 , the ratio of ethylene oxide to ethylene, is then small. This can be written

$$g = \frac{\log P_1/p_1}{p_2/p_1} = m - \frac{m^2 n}{2} \frac{p_2}{p_1} + \dots \quad (9)$$

The function g is plotted against p_2/p_1 , and the extrapolation back to $p_2/p_1 = 0$ gives m . Figure 9 illustrates this evaluation for the series S1. This estimation of m could be checked from the gradient of g when once n was evaluated, and in certain cases the next term of the expansion in equation (9) was considered.

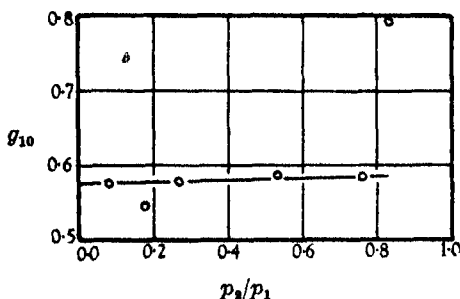


FIGURE 9. Evaluation of m .

The quantity $A_2 f_2(\theta)$ is obtained by estimating n , which may best be done from the maximum amount of ethylene oxide formed in the reaction. From equation (4) it can be shown that the maximum amount of the original ethylene which is converted to ethylene oxide is given by

$$\frac{p_2}{P_1} = \frac{1}{m} (1 - n)^{1/n-1}, \quad (10)$$

and using the value of m obtained above, n is found.

In this way, all the required quantities can be evaluated. The assumption that θ is constant will be least inaccurate at the beginning of a reaction, and hence $A_1 f_1(\theta)$ and $B_1 \phi_1(\theta)$ should be fairly accurate. But since $A_2 f_2(\theta)$ is derived from the maximum ethylene oxide when the reaction has gone to a considerable extent and θ will have altered, it will be much less accurate. Table 2 lists the values found.

In formulating the functions $f_1(\theta)$, etc., the difficulty arises that θ is not directly measurable in these experiments. It is certain, however, that θ will increase with increasing oxygen pressure when the ethylene pressure is kept constant (S1-3), and that it will decrease with increasing ethylene pressure when the oxygen pressure is constant (S4, 5, 7). This is shown by the figures for $A_0(1 - \theta)^2$, which suggest that θ is determined mainly by the oxygen/ethylene ratio.

In series S1-3 at constant ethylene pressure, when the oxygen pressure is increased fourfold, the values of $A_1 f_1(\theta)$ increase approximately twofold, and those of $B_1 \phi(\theta)$ approximately fourfold. It seems reasonable that in the conversion of ethylene to ethylene oxide, since one atom of oxygen is involved, the rate of reaction should be proportional to the first power of the concentration of adsorbed oxygen, i.e. $f_1(\theta) = \theta$. Then these results would indicate that $\phi_1(\theta) = \theta^2$, i.e. that the rate of the direct production of carbon dioxide is proportional to the square of the adsorbed oxygen concentration. If these conclusions are correct, then the ratio $[A_1 f_1(\theta)]^2 / B_1 \phi(\theta) = A_1^2 / B_1$ should be constant. The last column in table 2 lists these ratios, and it will be seen that with the exception of S5 and S7, these are reasonably constant. These discrepancies may be due to slight changes in the catalyst, which affect the calculations disproportionately. The series S1 and S7, which should be identical, appear from table 2 to have considerable differences, whereas it can be seen from figures 1 and 7 that the experimental differences are small.

TABLE 2. 263° C ^a

series	P_{Et} (mm.)	P_{O_2} (mm.)	m	n	max. EtO (%)	$A_1 f_1(\theta)$ $\times 10^3$	$B_1 \phi_1(\theta)$ $\times 10^3$	$A_1 f_1(\theta)$ $\times 10^3$	$A_0(1-\theta)^2$ $\times 10^3$	A_1^2/B_1 $\times 10^3$
S1	20	60	1.31	+0.02	28.3	3.65	1.15	4.7	2.1	11.5
S2	20	120	1.43	+0.31	31	5.0	2.2	5.0	1.42	11.4
S3	20	30	1.2	-1.7	18	2.4	0.4	7.6	2.67	14.4
S4	40	60	1.21	-0.9	21.5	2.53	0.53	5.8	2.4	12.1
S5	10	60	1.26	+0.18	32	5.8	1.5	6.0	1.52	22.3
S6	40	120	1.33	+0.31	33	4.1	1.4	3.8	2.14	11.9
S7	20	60	1.21	-0.07	29	4.8	1.0	6.2	2.05	23.0

It is possible in an approximate way to obtain values for the constants A_1 , B_1 and A_0 and for θ . Let the values of $A_1 \theta$ and $A_0(1-\theta)^2$ in table 2 be denoted by x and y respectively. Then these are connected by the relation

$$\sqrt{y} = \sqrt{A_0} - \frac{\sqrt{A_0}}{A_1} x. \quad (11)$$

A plot of x against \sqrt{y} is shown in figure 10; the data of S1-4 only are used, as the last three series of experiments do not appear to be so consistent. From the line obtained, values for A_0 and A_1 were deduced. The constant B_1 can be evaluated assuming an average value of 12.4×10^{-3} for the ratio A_1^2/B_1 (last column of table 2) in S1-4. The constants so found are

$$A_0 = 3.8 \times 10^{-2}, \quad A_1 = 13.0 \times 10^{-2}, \quad B_1 = 13.6 \times 10^{-2}.$$

With these constants, θ at the beginning of S1 is equal to 0.28. It may be mentioned that in the resistance experiments with a similar ethylene/oxygen ratio (46.3 mm. ethylene, 143 mm. oxygen) θ was estimated to be 0.30 at 301° C. It is obvious that the values assigned to these constants cannot be of a high degree of accuracy owing to their indirect method of derivation, but they served as the starting-point of the

exact solution of equations (1)–(3) by the differential analyser (this work will be published elsewhere), and were not very different from the values finally chosen by trial and error.

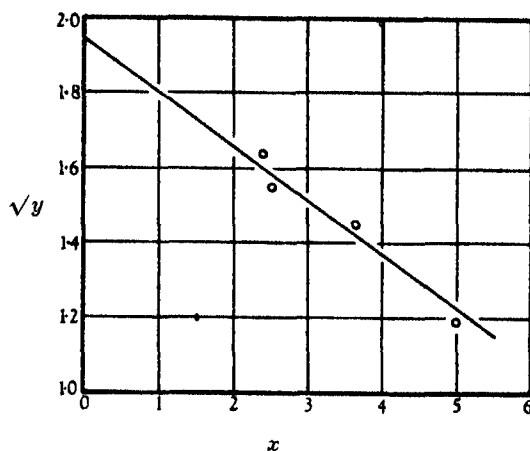


FIGURE 10. Evaluation of velocity constants A_0 and A_1 .

The oxidation of the ethylene oxide is controlled by the factor $A_2 f_2(\theta)$. The accuracy of measurement is much lower than that for the other factors, since it depends on (a) an accurate estimate of the maximum ethylene oxide production and (b) on the correct measurement of the quantity m . But the greatest error in the measurement of $A_2 f_2(\theta)$ is due to the assumption that θ is constant with time. Despite these difficulties, however, the absence of any trend in the values of $A_2 f_2(\theta)$ with variation in the ethylene and oxygen pressures suggests that this factor and therefore the rate of oxidation of ethylene oxide is independent of θ .

This can be seen more clearly by considering the effect of different kinetics on the maximum ethylene oxide formed. If the oxidation of ethylene oxide, and also the direct oxidation of ethylene to carbon dioxide, were first order with respect to θ , then the functions m and n would be independent of θ , and equations (4) and (10) would hold accurately independently of any variation in θ . In this case, it follows from equation (10) that, as the oxygen pressure and thus θ is increased, the maximum ethylene oxide produced would remain unaltered. The effect of the direct oxidation of ethylene to carbon dioxide, being second order with respect to θ , would be to cause the maximum to decrease slightly as the oxygen pressure was increased. The effect actually found was that the maximum increased as the oxygen pressure increased, and thus the oxidation of the ethylene oxide must be of a lower order than unity with respect to θ , and probably independent of θ .

Integration of kinetic equations

The most satisfactory confirmation of the kinetics deduced in this way, is by the integration of equations (1)–(3). Whilst this cannot be done analytically, the equations

have been solved by using the Bush differential analyser. This method only allows of solutions when the constants A_0 , A_1 , B_1 and A_2 are known numerically and a numerical value has to be chosen for the experimentally unobservable θ at the start of reaction. The details of the machine set-up for the differential analyser and the method of choosing θ will be published elsewhere.

The constants A_0 , etc., were estimated by trial and error, using the values given on p. 132 as guidance, and fitting the solutions to the series of experiments (S1) in which the initial ethylene and oxygen pressures were 20 and 60 mm. respectively. The solutions are shown in the solid curves of figures 1 and 7. The fact that such a good fit was obtainable with the experimental results is considerable confirmation of the kinetics assumed, but the real test was in the application of these solutions with the constants chosen by fitting to S1 and 7 to the other series of experiments in which different initial ethylene and oxygen pressures were used. The agreement is extremely satisfactory (except in the case of S3, where there seems to be some experimental error) and affords striking confirmation of the kinetics deduced. In particular, the agreement at the maximum of the ethylene oxide curves is a satisfactory proof that the rate of oxidation of ethylene oxide is independent of the adsorbed oxygen concentration.

Reactions between ethylene oxide and oxygen

It was thought that confirmation of the kinetics of the oxidation of ethylene oxide might be obtained from direct experiments between ethylene oxide and oxygen. The experimental details were the same as in the case of ethylene and the results are shown in table 3. The values of A_2 were obtained from the gradients of the straight-line graphs of $\log p_2/P_2$ against time; $A_0(1 - \theta)^2$ was similarly found from the $\log p_0/P_0$ graphs.

TABLE 3. $T = 263^\circ \text{C}$

series	p_{EtO} (mm.)	p_{O_2} (mm.)	A_2 $\times 10^3$
S 11	20	25	1.85
S 13	20	50	2.38
S 12	20	100	3.31

The results obtained here appear to be in conflict with the previous ones in that A_2 is not constant. This anomalous result is due to the fact that the oxidation of ethylene oxide goes by two stages, (a) isomerization to acetaldehyde, and (b) oxidation of the acetaldehyde; the latter step is very much faster than the former, and provided that the rate of adsorption of oxygen by the catalyst is sufficiently fast, no acetaldehyde accumulates during reaction, and the overall rate of oxidation of the ethylene oxide is that of the isomerization—a step which is independent of the oxygen concentration.

In the experiments with ethylene there was always a high oxygen/ethylene oxide ratio, and the rate of adsorption of oxygen by the catalyst was always ample to

oxidize the acetaldehyde; in the present series, however, this was not so, and since the analytical technique did not distinguish between ethylene oxide and acetaldehyde, the values of A_2 really refer to the production of carbon dioxide and water, and thus are dependent on the oxygen concentration. It will be noticed that the values of A_2 are less than that expected (4.0×10^{-2}), which agrees with this explanation.

Since the quantity of oxygen taken up by the catalyst is relatively small, the rate of uptake of oxygen by the catalyst should equal the rate of oxidation of the ethylene oxide, i.e.

$$A_0(1 - \theta)^2 p_0 = 2.5 A_2 p_2.$$

For there to be no acetaldehyde accumulating, $A_0 p_0$ must be greater than $2.5 A_2 p_2$, i.e. $p_0/p_2 > 2.5 A_2/A_0$. Assuming the values for A_0 and A_2 found to fit the ethylene oxidations ($A_0 = A_2 = 4.0 \times 10^{-2}$), this gives $p_0/p_2 > 2.5$. Of the above series of experiments, only S12 fulfils this condition, and the low values of A_2 found in S11 and 13 are due to the rate of uptake of oxygen by the catalyst being insufficient to oxidize all the acetaldehyde.

Another cause of the slower oxidation is the partial covering of the catalyst by a deposit formed from the ethylene oxide. In the previous paper it was shown that this deposit forms to a much greater extent with ethylene oxide than with ethylene. The formation of this deposit during the reaction between ethylene oxide and oxygen was confirmed by the fact that the increase in pressure of the system with time was always less than that calculated from analysis of the amount of carbon dioxide and water formed. That the substance going on to the catalyst was organic was shown by heating the used catalyst in oxygen when carbon dioxide and water were produced.

The experiments are thus not sufficient to determine the kinetics of this oxidation owing to the large amounts of oxygen required and more particularly to the organic deposit formed on the catalyst. It is more reliable to determine the kinetics from the experiments between ethylene and oxygen, since, in these, the rate of oxidation of the ethylene oxide is measured relative to the rate of production, and any deposit formed on the catalyst would affect both of these reactions equally.

Reaction mechanisms

The kinetics found in the present experiments agree perfectly with those expected from the results of the reactions between adsorbed oxygen and the various organic gases. There are two routes by which ethylene may react. On the one hand, it may take up a single adsorbed oxygen atom to form ethylene oxide; as would be expected the rate of this reaction is proportional to the first power of the concentration of oxygen adsorbed on the catalyst. On the other hand, the ethylene may be completely oxidized to carbon dioxide and water without first forming ethylene oxide; the rate of this reaction is here found to be proportional to the square of the adsorbed oxygen concentration, in agreement with the previous experiments with oxygen films. The first step in this process must therefore be reaction between an ethylene molecule and a pair of adsorbed oxygen atoms. Representations of the physical

processes involved in these two reactions of ethylene have been given in the previous paper.

The fact that agreement exists between the results obtained here and those found in the experiments with oxygen films is also a proof that neither homogeneous reactions nor reactions between adsorbed ethylene and gaseous oxygen take place to any appreciable extent. It is also possible to make an approximate correlation of reaction rates; the value for the velocity constant B_1 of the direct oxidation as estimated in the present experiments and used in the machine solutions was 16.7×10^{-2} , which is intermediate between the highest and lowest values found for B_1 in the experiments between ethylene and oxygen films (table 5, Part II).

In the previous paper it was shown that ethylene oxide could undergo a variety of reactions. It could (a) be isomerized to acetaldehyde and thence oxidized to carbon dioxide and water, (b) revert to ethylene and adsorbed oxygen, (c) be adsorbed on the catalyst and thence be further oxidized, or (d) possibly be oxidized directly to carbon dioxide and water.

The result found here, that the rate of oxidation of the ethylene oxide is independent of the concentration of adsorbed oxygen, fits in with the first of these processes, the isomerization to acetaldehyde, and indicates this as being the major reaction of the ethylene oxide. It is possible that some ethylene oxide is oxidized directly and not via acetaldehyde, but if this is so, the evidence of this and the previous papers is that only a small fraction of the ethylene oxide follows this path. Some ethylene oxide is adsorbed to form an organic deposit covering up part of the catalyst; this is slowly oxidized to carbon dioxide and water, but, as was shown in the previous paper, the quantity of ethylene oxide following this path is small.

In the experiments made with ethylene oxide alone in contact with the catalyst, the indications were that the reaction back to ethylene and adsorbed oxygen was relatively important. However, in the series of reactions between ethylene oxide and oxygen, no ethylene was ever detected in the products. This surprising result can be interpreted if it is assumed that there are two types of silver surface concerned in the adsorption of oxygen; one, amounting to perhaps as much as 20 % of the surface, which adsorbs oxygen more strongly than the other. It is this part of the surface which, because of its great affinity for oxygen, brings about the reversion of ethylene oxide to ethylene. When ethylene oxide is admitted to an oxygen-covered catalyst, the acetaldehyde formed causes the removal of the strongly bound oxygen, after which ethylene can be produced. But when oxygen is present in the gas phase, this strongly adsorbing part of the surface remains covered with oxygen and no ethylene can be produced. Evidence that the silver surface is not uniform has been found from experiments on the adsorption of oxygen; it was found that when an oxygen-covered surface was allowed to evaporate into a high vacuum, a large part of the oxygen was removed, but a certain fraction remained even after long pumping; this fraction could, however, be removed by chemical means, e.g. cleaning up by ethylene or hydrogen.

Effect of temperature

The effect of temperature on the oxidation of ethylene can be seen from figures 11 and 12, which are to be compared with figures 1 and 7; in all these experiments, the initial mixture was 20 mm. ethylene and 60 mm. oxygen. Apart from the general increase in reaction rates with temperature, the most noteworthy features shown by these curves are (a) the increase in the maximum amount of ethylene oxide produced, and (b) the slight increase in the direct reaction to carbon dioxide. It will be shown

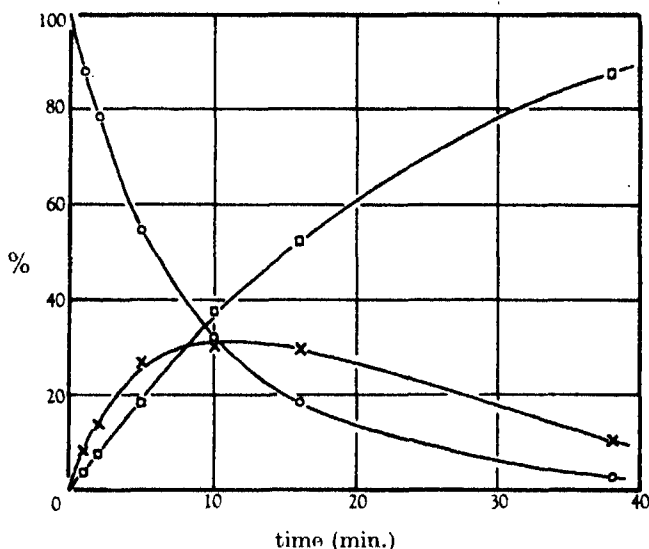


FIGURE 11. Course of reaction between ethylene and oxygen at 285° C.

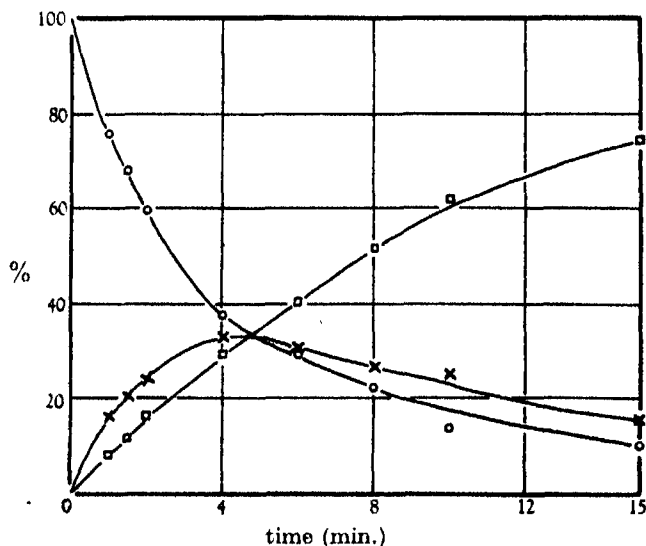


FIGURE 12. Course of reaction between ethylene and oxygen at 302° C.

later that both of these effects are due to an increase with temperature in the concentration of adsorbed oxygen, rather than to changes in the order of the velocity constants.

The velocity factors deduced from these experiments are given in table 4. It is not possible directly to obtain the energies of activation from these values as θ is not measurable. However, approximate values for these energies can be estimated by combining the present data with that obtained elsewhere.

TABLE 4. EFFECT OF TEMPERATURE

series	$T^\circ\text{C}$	m	n	max. EtO (%)	$A_1\theta$ $\times 10^2$	$B_1\theta^2$ $\times 10^2$	A_2 $\times 10^2$	$A_0(1-\theta)^2$ $\times 10^2$	A_1^2/B_1 $\times 10^2$
S 1	263	1.31	+0.02	28.3	3.65	1.15	4.7	2.1	11.5
S 8	285	1.40	+0.30	31	9.0	3.6	8.8	5.3	22.5
S 9	302	1.36	+0.35	33	20.4	7.3	18.0	10.3	57

The experiments on the clean-up of oxygen films by ethylene have provided directly values for the constant B_1 from which an energy of activation E_{B_1} of 10 ± 3 kcal. for the direct oxidation of ethylene to carbon dioxide was obtained (Part II, p. 119). Experiments using the resistance effect gave for the clean-up of oxygen by ethylene an energy of activation of 10 kcal.; although the resistance effect is somewhat complex, this can be taken as confirmation of E_{B_1} .

By using the figures of table 4 and plotting $\log A_1\theta$ against the reciprocal of the absolute temperature, the apparent energy of activation for the production of ethylene oxide was found to be 26 ± 2 kcal., in agreement with the value of 27 kcal. found from flow experiments (Part I, p. 99). Similarly, the apparent energy of activation of the direct oxidation (from $B_1\theta^2$) is 29.5 ± 1.5 kcal. A combination of these two values, or alternatively, a plot of $\log A_1^2/B_1$ taken from the last column of table 4, gives

$$2E_{A_1} - E_{B_1} = 22 \pm 5 \text{ kcal.}$$

Using the value above for E_{B_1} , an estimate of 16 ± 4 kcal. follows for A_1 .

The velocity constant A_2 for the oxidation of ethylene oxide can be obtained directly from table 4. As explained before, these figures are bound to be inaccurate, but from them an approximate estimate of the energy of activation of this reaction can be made. This comes to 20 ± 5 kcal. and is the energy to be associated with the isomerization of ethylene oxide to acetaldehyde.

The apparent energy of activation for the adsorption of oxygen by the catalyst, taken from the values of $A_0(1-\theta)^2$, was 24.8 kcal. Since the apparent energies of activation for ethylene oxide production, and the direct oxidation to carbon dioxide are both greater than the true energies of activation associated with these steps, it follows that as the temperature is raised, θ increases. So for the uptake of oxygen, one would expect the true energy of activation associated with the velocity constant A_0 to be greater than 25 kcal. Measurements on the resistance effect indicated a value of 27 kcal.

These figures can be summarized:

oxygen adsorption	$E_{A_0} > 25 \text{ kcal.}$
ethylene oxide production	$E_{A_1} \quad 16 \pm 4 \text{ kcal.}$
direct CO_2 production	$E_{B_1} \quad 10 \pm 3 \text{ kcal.}$
ethylene oxide oxidation	$E_{A_2} \quad 20 \pm 5 \text{ kcal.}$

From these figures it follows that as the temperature is raised the value of θ during reaction will increase until adsorption becomes rapid relative to the other reactions, when θ will approximate to the value obtaining in the absence of ethylene. This latter value decreases with increasing temperature, so that θ will have a maximum value. Corresponding to this, the maximum in the curve of ethylene oxide against time will reach an upper value (estimated at about 35–40 %) as the temperature is raised and decrease rapidly thereafter, since E_{A_2} is greater than E_{A_1} . Similarly, the fraction of ethylene oxidized directly to carbon dioxide will increase at first as the temperature is raised, owing to the increase in θ , but will subsequently decrease rapidly owing to the fall in θ and the low value of E_{B_1} .

An approximate measurement of θ

An attempt was made to measure the concentration of adsorbed oxygen on the catalyst during reaction by pumping out the contents of the reaction vessel and cleaning up the oxygen with hydrogen. The catalyst, which in the experiments immediately preceding had been in contact with ethylene oxide, was cleaned as far as possible from organic deposit by heating at 330°C in oxygen for 14 hr. The temperature was reduced to 263°C after pumping out the reaction vessel, and the catalyst was left in contact with 50 mm. of oxygen for half an hour. In order that θ might settle down, three reactions with 20 mm. ethylene and 60 mm. oxygen were run, each to the point where ethylene oxide was a maximum. After the third run, the reaction vessel was pumped out for 1 min., and 60 mm. of hydrogen was admitted. After 22 min. this was pumped through the liquid-air trap in the analyser and the condensable products analysed. A small quantity of carbon dioxide was found, equal to 0.6 mm., together with 6.2 mm. of water (all quantities given in equivalent pressures in the reaction vessel). It is thought that the carbon dioxide found here comes from the organic deposit formed during the ethylene-oxygen reaction and displaced by the hydrogen, and so the oxygen in it is not counted. From the quantity of water found, the amount of oxygen adsorbed during reaction is equivalent to 3.1 mm. This figure may be slightly high if some water left over from the ethylene-oxygen reaction and adsorbed on the catalyst is displaced by the hydrogen.

It is estimated (unpublished work) that the quantity of oxygen which would cover the catalyst at saturation is equivalent to about 10 mm. pressure in the reaction vessel. Consequently, the value of θ indicated by the above experiment is 0.31. Considering the difficulties of this experiment, this compares not unfavour-

ably with the value of 0.22 which fitted the machine solutions for the rate equations under similar conditions. In the presence of oxygen alone, θ would have had a value of about 0.6.

CONCLUSIONS

In the previous papers the complex of reactions that occur when ethylene and oxygen interact in the presence of a silver catalyst was dissected and the individual reactions examined. The present work has shown that when the overall reaction is considered, it is in agreement with what is expected when the individual steps are put together. The first step in the reaction is the adsorption by the catalyst of oxygen in the form of atoms. Ethylene reacts either with one atom to form ethylene oxide, or with a pair of atoms to form some intermediate which is then rapidly oxidized further to carbon dioxide and water. Ethylene oxide in turn can be oxidized; it first undergoes isomerization to acetaldehyde which is rapidly oxidized to carbon dioxide and water. In addition, there are minor reactions involving the formation and oxidation of an organic deposit on the catalyst, mainly produced from ethylene oxide.

The correlation between reaction rates found here and those obtained in the experiments with oxygen films throws some light on the problem of whether all or only a small fraction of the catalyst surface takes part in reaction. The velocity constant B_1 of the direct oxidation to carbon dioxide has approximately the same value as that deduced from experiments on the cleaning up of oxygen films by ethylene. Consequently, a large part, and possibly the whole, of the catalyst surface must be engaged in reaction.

It will be noted that in the mechanisms found here, reaction occurs between one component (oxygen) which is chemisorbed on the catalyst and the other component (ethylene) which is either not adsorbed at all, or only weakly adsorbed by van der Waals forces. This general type of mechanism has been found previously in the case of reactions between ethylene and hydrogen on a nickel catalyst (Twigg & Rideal 1939); in that instance it was the ethylene which was chemisorbed and the hydrogen which was loosely held to the catalyst by molecular forces. It seems possible that this type of mechanism, where one component is chemisorbed and the other is not, may be a general one in heterogeneous catalysis. The more usual view of heterogeneous catalysis as reaction in a two-dimensional film (Schwab 1937) appears to be inherently less probable; it requires the simultaneous condition that both reactants are adsorbed, strongly enough to be in sufficient concentration, but not so strongly that they cannot be detached from the surface. Whereas, on the view proposed here, only one reactant need fulfil these conditions; and, indeed, the fact that silver is a good oxidation catalyst for a large variety of organic substances is due to the moderate adsorption of the oxygen and does not depend on the ability to adsorb the organic material.

It would be possible to express the results of the kinetic experiments of this paper by means of the formula $p_{\text{H}_2}^m p_{\text{O}_2}^n$, for the rate. If this is attempted the results do not

fit very well, but give approximately $m = 0.45$, $n = 0.55$ for ethylene oxide production, and $m = 0.3$, $n = 1.1$ for the direct oxidation of ethylene to carbon dioxide, figures which might suggest that ethylene was moderately strongly adsorbed and oxygen weakly adsorbed as atoms; in addition, the different values of m for the two reactions might suggest two types of surface for these reactions. This is in direct contradiction to the interpretation put forward in these papers, an interpretation which is amply supported by experiments other than purely measurements of rates of reaction. Thus it would appear that in investigating the mechanism of catalytic reactions there may be some danger in relying solely on correlations between reaction rate and concentrations of reactants.

It is with pleasure that I record my thanks to Professor E. K. Rideal for many stimulating discussions and his continued interest in this work. I am also indebted to the Distillers Company Limited for generous financial assistance.

REFERENCES

- Schwab 1937 *Catalysis*, Chap. 12. London: Macmillan and Co.
Twigg & Rideal 1939 *Proc. Roy. Soc. A*, 171, 55.

CORRIGENDUM

Proc. Roy. Soc. A, volume 185, number 1003, page 443, figure 4

Ordinates of the figure: add one pH unit to each
number on the axis—now to read 11 to 15.

Address of the President
Sir Robert Robinson, at the
Anniversary Meeting, 30 November 1946

The grievous losses sustained by the Society during the past year have just been announced. Our *Obituary Notices* are the appropriate medium for the appreciation of the services to science and the Society of the deceased Fellows but, before following the example set by Sir Henry Dale last year and proceeding to the presentation of Medals, I will make a few very brief comments. First that Sir James Hopwood Jeans, O.M. was awarded a Royal Medal in 1919 for his distinguished contributions to astronomy and he was the 'A' Secretary for the next ten years. We pay a tribute to the memory of an active and conscientious former officer of the Society whose eminence in the scientific field has received world-wide recognition.

Professor Percy Faraday Frankland, C.B.E., was elected a Fellow sixty-five years ago. He was awarded the Davy Medal in 1919 and served as Vice-President in 1917-18.

During the past year Council nominated the late John Maynard Keynes, first Baron of Tilton, for election as a Fellow of the Society under Statute 12. His untimely death robbed the Society of the opportunity to add the lustre of his name to the records.

Mr H. G. Wells was a friend of science whose imaginative writings and popular expositions have had an influence which it would be hard to overestimate. Though he was not a Fellow of the Society I think it right that we should 'praise him as a famous man' and add our voices to the general acclaim of his achievement.

Awards of Medals, 1946

The COPLEY MEDAL is awarded to Professor EDGAR DOUGLAS ADRIAN, O.M., for his outstanding contributions to nerve physiology.

During the last thirty years Professor Adrian has been engaged on a series of systematic investigations of the essential functions of the nervous system which have been extended from a study of the activity of single nerve fibres to the reaction of the cortex of the forebrain of man to impulses that reach it from the periphery. The advance of our knowledge of the working of the nervous system is largely the result of his researches into the nature of the fundamental process of individual cells and combinations of cells.

His early work with Keith Lucas provided important observations on conduction by nerve fibres and on the reactions of muscles. This was followed by a series of independent researches by the combination of a valve amplifier with a capillary electrometer which made possible an analysis of the behaviour of individual sensory receptors and of single motor units. Among many important discoveries

these investigations revealed how the frequency of impulses conveyed by each fibre is used in the central nervous system to signal the intensity of peripheral and central events.

By the same methods he undertook a detailed analysis of the activity of many types of sense organs and of simpler reflex actions in terms of the activity of single nerve fibres, and in some cases demonstrated that the same principles underlie all nervous activity throughout the animal kingdom. Other investigations dealt with the nature of the fundamental process in nerve cells and in synaptic regions of the central nervous system.

During the past ten years he has been mainly concerned with the interpretation of the potential waves in the cortex of the forebrain. Hans Berger had drawn attention to the existence of these in man, but the subject was neglected until Adrian and Matthews reinvestigated it. Adrian's subsequent studies included, in the first place, an examination of the electrical activity of the brain and its reaction to messages from the periphery, and in the second place a mapping out of the regions of the cortex which serve as receiving centres for such messages. His aim has been to analyse these phenomena in terms of activity of simple nervous units, and the results of his work are the basis of the subsequent development of electroencephalography, which has attained an important place in both physiological and clinical investigations.

By his researches on the exposed brains of animals he determined the laws of spread of activity in the cortex, its reactions to natural and artificial stimuli that reach it, and showed that the interaction between a local excitation and the background of spontaneous activity is the essential feature of a cortical response.

By a study of the comparative physiology of the sensory areas of the brain he has also shown how their development and to some extent their reactions to peripheral stimuli depend on the structure and mode of life of the animal.

After determining the representation in the brain of receiving stations for superficial and proprioceptive stimuli he investigated that of vision and hearing. He has even succeeded in demonstrating the different features of impulses that reach the visual cortex from the rods and cones of the retina; he has also dealt with the distribution and significance of certain non-sensory afferent impulses, as those that reach the cerebellum.

Adrian has blazed many trails in his exploration of the territory of nerve physiology. It is certain that for many years to come his lead will be followed and the new knowledge will be consolidated along the lines of his pioneering work.

The RUMFORD MEDAL is awarded to Sir ALFRED CHARLES GLYN EGERTON for his distinguished researches on combustion.

The Rumford Medal founded in 1800 is awarded once every second year 'to the author of the most important discovery or useful improvement which shall be made and published by printing or in any way made known to the public in any part of Europe during the preceding two years on Heat or on Light, the preference

always being given to such discoveries as, in the opinion of the President and Council of the Royal Society, tend most to promote the good of mankind'.

Sir Alfred Egerton admirably fulfils the requirements of these terms of award. He is a physical chemist whose researches have always been directed towards the application of physico-chemical principles to the process of combustion of hydrocarbons in all its ramifications. For some time the approach to combustion problems has been empirical because there was no satisfactory physico-chemical basis of the theoretical or practical aspects to make further significant progress possible. This background has now been partly provided by Egerton who was one of the first to see clearly how necessary it was to apply the new conceptions of combustion to the complex processes occurring under the conditions obtaining in internal combustion engines. One of the great obstacles to achieving greater efficiency is the difficulty of preventing premature detonation. This phenomenon is essentially a chemical one in the sense that organic peroxides, produced during combustion, are known to be responsible for the pre-ignition. Thus the chemical behaviour of peroxides might provide a key to the solution of the problem and much of Egerton's work has been devoted to this enquiry. The investigation involved the elaboration of special physical techniques since ordinary chemical methods were inapplicable to this type of research.

During the war Egerton has directed his attention to the vital problem of ensuring that combustion appliances should be devised and operated with the maximum possible efficiency. This can only be achieved by a thorough scientific analysis, hitherto lacking, mainly because the problem had hardly been considered worthy of serious study. The result of his labours cannot fail to be of great benefit to the country during the period of very low fuel production and will lead to considerable economies under all circumstances.

The characteristic of Egerton's work has been the application of modern physico-chemical methods to current scientific and technical problems of great moment, combined with experimental researches developed with great ingenuity.

A ROYAL MEDAL is awarded to Sir WILLIAM LAWRENCE BRAGG for his investigations of the structure of solids.

The diffraction of X-rays by crystals was observed in 1912 by Laue, Friedrich and Knipping, but the pioneers of the present-day development are the late Sir William Bragg and his son Sir Lawrence. It was W. L. Bragg who formulated the law $n = 2d/\sin \theta$ that is now so familiar in all studies concerned with the structure of molecules and their states of aggregation. Soon after Laue's discovery there followed, from father and son, a series of papers on the phenomena of X-ray 'reflexion' on the one hand and the determination of fundamental crystal structures on the other, the far-reaching consequences of which could hardly have been foreseen, even by their authors. At the present time crystal analysis by X-rays is an established technique, a sharp tool of research that lays bare the complexities of organic and mineral matter alike.

The inspiration and genius of Bragg are seen in so many of the modern developments of X-ray diffraction to structural analysis that it is possible to select only a few outstanding illustrations. His principal interest has always lain in the interpretation of diffraction phenomena with the view of making the actual methods of analysis more precise, more simple, and more extended. In developing such methods he and his collaborators have elucidated the atomic arrangement in a great number of fundamental types of inorganic crystal structures. Chief among these are those of the diamond and the elementary salts and oxides, in the study of which the subject found its first beginnings. After these perhaps his greatest analytical success is shown in the field of the silicates. A chemical riddle has been transformed into a system of simple and elegant architecture. He has also contributed greatly to our knowledge of the structure of metals and alloys and their phase changes, and of the relations between their physical properties and atomic arrangement in the crystalline state. Latterly he has brought to a still clearer focus the concept of X-ray diffraction as a branch of optics and has thus initiated methods that have already gone far towards replacing the earlier laborious calculations by rapid devices based on the analogy of the diffraction of visible light.

The implications and applications of the principles and methods of X-ray spectroscopy and X-ray structure analysis are one of the wonders of modern science, and with this manifold triumph the name of Sir Lawrence Bragg is inseparably associated.

A ROYAL MEDAL is awarded to Dr CYRIL DEAN DARLINGTON in recognition of his distinguished contributions to cytology.

The importance of Darlington's work lies not so much in the discovery of isolated new phenomena—although he has discovered many of these—but rather in the achievement of a synthesis which brings together a highly diversified body of apparently disconnected facts into an integrated system.

Darlington's first major achievement was the clarification of the relations between the two main forms of nuclear division—mitosis and meiosis. Out of the confused mass of available observations, he singled out as fundamentally significant two facts, first, that chromonemata attract one another specifically, by an attraction which is satisfied when two similar threads are associated; and second, that in the earliest stage of mitosis each chromosome is already split into two halves, while at the beginning of meiosis they are still single. On the basis of these two facts, he showed that the relation between the two forms of division could be understood as the result of a temporal shift in the operation of a single physiological process. His so-called 'precocity' theory of meiosis was then supported by a whole series of new observations, in which the resources of comparative study and of new techniques were used for the specific purpose of obtaining answers to critically formulated questions.

Starting from the basis of the relation between the two major forms of nuclear division, Darlington has pursued his enquiry in two directions. On the one hand,

he has accumulated a very large body of facts concerning the detailed mechanics of cell division in many different groups of plants and animals. The comparative method enabled him to reach important new conclusions as to the mechanism of crossing-over, the cycles of spiralization and contraction of chromosomes, the nature of the mitotic spindle and the forces exerted by it and within it, the role of the centromeres and so on. These results have laid a broad foundation of observation and deduction which appears, for the first time, firm enough to bear a superstructure of physico-chemical interpretation. Proceeding in quite another direction, Darlington discussed the implication of his cytological ideas on evolutionary theory. The existence of the mitotic and meiotic modes of division had been explained as the result of different modalities in the application of a single set of physiological principles; Darlington showed that further, slighter modifications could produce many of the widely diverse series of reproductive mechanisms met with in the animal and plant kingdoms. He emphasized the fact that the mechanism of evolution is itself subject to evolutionary changes.

More recently Darlington's work has led him to the investigation of the general problems of gene action, of the physiological action of the two major types of nucleic acid, and the relation between the gene and other similar bodies in the cytoplasm. Darlington was one of the first to enter this highly speculative field and he has contributed not only his full quota of stimulating speculation, but also a large share of the still scanty facts. It is not too much to say that Darlington's results and theories are recognized as the basis of modern nuclear cytology.

The DAVY MEDAL is awarded to Professor CHRISTOPHER KELK INGOLD for his outstanding researches in physico-organic chemistry.

Progress in one of the most active fields of chemical science during the present century has resulted from attempts to elucidate the detailed mechanism of organic reactions in terms of modern physical concepts. Throughout this development Ingold's contributions are especially distinguished. Possessing detailed knowledge and understanding of both the physical and organic branches of the science he has been in a position to effect the synthesis of the two modes of approach without which a successful attack on the difficult, yet fundamental, problems involved could not be achieved. It is not possible in short compass even to outline the range of investigations with which Ingold has been concerned, but brief mention may be made of the work on stereochemistry dealing with ring strain and the effect of *gem* dimethyl groups on the valency angles of carbon. A further application of underlying physical principles is evident in his investigations of tautomerism in triad systems, and in the development of our ideas on ring-chain tautomerism. This work led on to more general studies of the mechanism of reactions, including the difficult question of substitution in the benzene ring, in addition to the ordinary reactions of organic chemistry, such as hydrolysis and substitution, which despite their apparent simplicity have proved to be complicated and difficult to interpret. The success which Ingold has achieved in interpreting these phenomena in terms

of the electron theory of valency is striking but in addition he has played the most prominent part in the experimental investigations which have led to our present knowledge of the kinetics and mechanism of organic chemical reactions. Ingold always has been interested in the elucidation of the course of chemical change by application of physico-chemical methods based on reaction velocities and in this field may be cited the extensive work on the mechanism of substitution at an aliphatic carbon atom, leading to the recognition of the uni- and bi-molecular processes, by means of which so much has been done to solve the difficult problems raised by the Walden inversion and the phenomena of racemization. In these intractable regions the contribution of Ingold and his flourishing school are of fundamental importance. Another aspect of his work involves a still deeper concern with physical principles as applied to organic chemical problems. His interest in the chemistry of benzene has led him to investigate in the fullest detail, using infra-red and Raman spectra, and indeed all available physical methods of approach, the fine structure of the benzene molecule. In order to provide the necessary data it was necessary to devise methods for the preparation of the various deuterium substituted benzenes—no mean feat of organic chemistry in itself—and the interpretation of the experimental results in terms of quantum mechanical principles has recently been published in an issue of the *Journal of the Chemical Society*, which he monopolized. Although his theoretical contributions have attracted more attention, the originality of his experimental technique is equally noteworthy and his happy selection of crucial tests amounts to genius.

The DARWIN MEDAL is awarded to Sir D'ARCY WENTWORTH THOMPSON, in recognition of his distinction as a zoologist.

Sir D'Arcy Thompson is now in his 62nd year as a Professor of Biology and Natural History. He published his first scientific paper in 1879. His most distinguished work, *On growth and form*, appeared in 1917 and was republished in a new and enlarged edition in 1942. He is still writing, but mainly in the field of the classics where he is a considerable scholar, and a great authority on all animals that have appeared in classical texts.

D'Arcy Thompson's scientific work ranges over a wide field of general zoology and marine biology. He is an expert on the subject of fisheries, and for a considerable period did tireless work both for the Conseil International pour l'Exploration de la Mer and the Fishery Board for Scotland, carrying out hydrographical observations and being responsible for a great deal of fishery statistics. In the main, however, D'Arcy Thompson's scientific reputation rests on his work on growth, and the dimensional relationships of animal forms. The better part of the foundation of modern research into these subjects is his demonstration of methods by which the shape of the living organism can be brought into the field of controlled mathematical enquiry.

D'Arcy Thompson's work springs essentially from an enquiry into the relationships of animal forms, and from an attempt to introduce a degree of mathematical

precision into the otherwise purely descriptive language of systematic evolution. He was able to show, for example, that the evolution of one form from another could often be illuminated by the use of Cartesian transformations. By making clear the formal unity and coherence in the relationship of animals which apparently differ in a multitude of ways, his studies made possible the quantitative demonstration of steps in the evolution of different forms, and more so, the orderly process of change in the development of the same form. In his own words, growth can be studied as a systematic deformation of form at an earlier stage. He showed, for example, that relative growth rates in different parts of the body are distributed according to an ordered system of growth-gradients. This concept can be applied to certain types of evolutionary transformation, since it helps to explain how a single genetic change can automatically affect both the size and the growth interrelations of several organs. The development and illustration of the theory of allometry is another extension of D'Arcy Thompson's ideas.

The wide variety of problems to the solution of which D'Arcy Thompson has opened the door is well indicated in the 'Festschrift' presented to him last year. In introducing his classic *On growth and form*, D'Arcy Thompson declared that it required no preface, since it was all preface. His elaboration of the subject covers so wide a field, however, that, until such time as some different and all-embracing set of general propositions is put forward to take the place of those he propounded, individual contributions to study of growth and bodily transformation must necessarily represent isolated developments of the structure which he has presented to us. D'Arcy Thompson's work will always be regarded as a necessary step in the development of biological knowledge.

The SYLVESTER MEDAL is awarded to Professor GEORGE NEVILLE WATSON in recognition of his distinguished contributions to mathematical analysis.

Watson is a mathematician of outstanding perseverance and analytical skill. For forty years he has devoted his energies to pure mathematics, and has made many important and exhaustive contributions, particularly in the field of analysis. The most important researches of Watson's earlier period are those on asymptotic expansions: his great memoir 'A theory of asymptotic series' appeared in the *Philosophical Transactions* of 1911 and was followed by a stream of other writings dealing with the characteristics and transformations of these series, and with their applications to several well-known functions of importance in mathematical physics. This group of discoveries has enriched the region of mathematics in which Stirling was the pioneer two centuries ago, and where the methods of approximation are reduced to a precise science. These include the method of 'steepest descent' and any account of asymptotic series to-day must be based to a great extent upon Watson's discoveries.

About that time many of the ablest pure mathematicians were trying to sum a difficult oscillating series involving Bessel and Legendre functions, which had presented itself in the theory of the diffraction of wireless waves round the earth.

Watson solved the problem by a new method (1918), and went on to study the more difficult case of the transmission of electric waves when it was assumed that the earth is surrounded by a concentric conducting layer, as suggested by Heaviside. Watson's powerful analysis made possible a great advance in the physical theory.

Another example of his capacity for providing a brilliant solution of a problem which had been attempted by many of his predecessors is furnished by his paper on the Rogers-Ramanujan identities. Following this came his work on general transforms in which he solved a problem which many celebrated mathematicians had attempted without success. This is probably Watson's greatest achievement, and ranks as one of the most important contributions to the subject made in recent years. In addition it has had the great merit of inspiring a large amount of work by other mathematicians. Among the more notable papers of the next years, were those on 'Generating functions of class-numbers', on Ramanujan's continued fraction, and one that gave the proof of Ramanujan's assertion about the number 691 which occurs in 'almost all' the terms of a certain infinite product when expressed as a series. The series of papers on singular moduli, during the period 1932-36, are deservedly celebrated.

His great book on Bessel functions is perhaps the most impressive single work that has ever been written on the analysis of functions. In collaborating with Professor E. T. Whittaker in the second and later editions of *Modern analysis*, he has shown the same breadth and power and has influenced the course of higher analytical mathematical teaching throughout the world.

THE HUGHES MEDAL is awarded to Professor JOHN TURTON RANDALL in recognition of his distinguished contributions to applied physics, and especially of his development of the magnetron.

In 1940 Randall, while working in the laboratory of Professor Oliphant, at Birmingham, agreed to join H. A. H. Boot in an endeavour to utilize the magnetron principle for the production of electromagnetic waves of frequency greater than 3000 Mcyc. Previous work in the laboratory had shown that satisfactory circuits for these wave-lengths must be an integral part of the internal structure of the valve itself. Randall and Boot together put forward the suggestions that such a circuit, for a multi-segment magnetron, should consist of a revolver-like arrangement of holes, spaced evenly about a circle, each hole communicating by means of a slot with a central cavity in which the cathode was mounted. The first trials with demountable valves using tungsten cathodes were immediately successful, and it was shown that the suggested form of valve can generate continuously oscillations of the required wave-length.

With the assistance of S. M. Duke, Randall and Boot were able to develop methods of construction of the magnetron which enabled oxide-coated cathodes to be used, and which therefore gave high powers when the valve was subjected to pulsed operation. Empirical investigation fixed the best coupling arrangement by which

the power could be fed into an external circuit. Careful investigation of the operation of a valve showed that it was subject to sudden changes of wave-length, a condition which limited its applicability to Service equipment. This difficulty was overcome by the 'strapping' methods developed by Dr J. Sayers.

There is little doubt that the magnetron valve was the prime factor in the improvements made in Radar during the war, and Randall deserves a very large share of the credit for this development.

He contributed also to the problem of crystal detection of centimetre waves.

Randall's studies of fluorescence and phosphorescence were of a high standard and his careful and painstaking experimental work did much to establish on a firm basis the theories of semi-conductors developed by Wilson and others, and especially the assumptions about the existence of 'electron traps'. His work on practicable phosphors has been of importance in the development of fluorescent lamps, and of the screens of cathode-ray tubes.

Randall has also made contributions to the X-ray investigation of the structures of glasses and of liquids, and he developed satisfactory forms of oxide cathodes for the fluorescent lamps.

I am greatly obliged to colleagues in the Council for their assistance in the preparation of notes on the Medallists.

It would certainly be convenient for me, but neither fitting nor necessary, to traverse from this Chair the ground covered by the Report of Council for the past year. His Majesty the King, accompanied by Her Majesty the Queen, opened the British Empire Scientific Conference on 17 June and each of the Delegates from the Dominions, the Colonies, and the United Kingdom was presented to their Majesties.

The Conference, and the Official Conference that followed, provided the opportunity for a veritable ferment of discussion and many wise and timely suggestions were made.

His Majesty the King drew attention to the desirability of holding similar conferences in the future in other cities of the Empire. Fellows will be glad to know that the Society is represented on the continuing body which will watch the progress of implementation of the resolutions that emerged from the Conferences.

The culmination of our activities was the tribute, I think we may claim, the worthy tribute, paid to the greatest of natural philosophers in our celebrations of the Isaac Newton Tercentenary.

The resounding success of these memorable occasions was made possible by the hard work and devotion of a few, by the willing co-operation of many, and by generous financial assistance. To all, the Officers, Council and Fellows of the Society are most deeply indebted. Many of our distinguished foreign guests have written letters expressing their pleasure in the visit and their satisfaction in meeting us in a truly international gathering. It was most unfortunate that exigencies of space, and other matters of which you will be cognisant, imposed a limitation on

the desired full participation of all Fellows in these Conferences. The necessity for appointed delegations from the United Kingdom was realized from the outset to be a necessary corollary, under present conditions, of the invitation to a large number of guests from abroad.

Professor Hill has served the Society as Biological Secretary for ten years and now as Foreign Secretary during the past year, always with the greatest zeal and efficiency. We shall sorely miss his wisdom in counsel, his occasionally puckish delight in giving the new idea a chance, and our ability to draw so often in the future as in the past on his unequalled store of knowledge of scientific matters and scientific men. We wish him all happiness and success in the experimental enquiries to which he now proposes to devote himself. You have just appointed his successor in the Foreign Secretaryship. The Society is indeed fortunate, that Professor Adrian, after a little natural hesitation, accepted the nomination of Council.

His world-wide fame as an investigator and discoverer and his extensive knowledge of scientific affairs overseas render the appointment peculiarly appropriate and one that will maintain the reputation, and sustain the prestige, of the Society in foreign lands. But the Foreign Secretary, as an Officer and member of Council, shares our general work and responsibilities. We look forward with keen anticipatory pleasure to his co-operation, and assure him of ours.

The Council has received with great regret the resignation of Mr Griffith Davies as Assistant Secretary. He was appointed in September 1937 and though I have only had one year's experience of his willing and able help, it has been amply sufficient to enable me to realize the extent of the loss that the Society has sustained. Mr Davies has become an authority on the records and history of the Society and is prepared, indeed is anxious, to continue his studies in those directions. Council has appointed him a member of the Library Committee and, subject to the concurrence of that Committee, he will be Chairman of a sub-committee charged with the duty of preparing for the celebration of the Tercentenary of the Society.

The inspired address that Sir Henry Dale delivered a year ago is fresh in our memory and his thesis that scientific knowledge should be a free gift to mankind found a ready response in all our hearts.

The warning of the possible danger to scientific ideals and integrity inherent in the conception of 'total war' must be heeded. We are, however, presented with some difficult practical problems that call for immediate solution and some of us may be on the horns of a dilemma. It is analogous to that which faces the leaders of religion in times of war and it was clearly felt quite recently in Parliament during the debate on conscription. This dilemma can only be resolved by real friendship and concord among the nations and is ever present in an armed peace. A definition courts inaccuracy and undue simplification but, very roughly, it is a question of a conflict between our ideals and conception of service to humanity on the one hand, and our duty as citizens and units of a democratic community on

the other hand. As Sir Henry rightly observed, the individual scientist must obey the dictates of his conscience.

Opinion is in a formative condition and before proceeding to the rather more congenial task of discussing certain recent scientific advances, I will offer some propositions for consideration. I need not remind you that the Royal Society as a body has no views on political questions, and therefore I speak for myself alone.

In the first place all scientists should strive to promote international amity and the outlawry of all methods of warfare which by their nature involve 'total war'. These include aerial bombardment of cities, chemical warfare, biological warfare, the use of atom bombs, and any similar devilries that may be devised. This is not an empty dream because the universal brotherhood of scientists has a real existence. Understanding between the nations will resolve such dubieties and what follows is set down on the assumption that agreement will unhappily not be reached, or will be inadequate to justify a relaxation of defensive measures.

My second point is that the clear-cut distinction between war and peace is fallacious in the present connexion. Under modern conditions a nation must prepare, in peacetime, its defences against a possible attack.

Thirdly, it is inconsistent to praise our scientists for their outstanding contributions to the war effort and at the same time to suggest that they offend against our ethical code if they serve the country in a similar fashion during an uneasy peace. It is useless to attempt to disguise the fact that such service implies some sacrifice of freedom. During the war the scientific effort was nation-wide and control extended to many university departments. Nevertheless the universities have preserved intact their precious liberty of action, and I see no signs of any attempt to curtail it. Surely this suggests a feasible line of demarcation in that extra-mural contracts, placed by Service Departments with the universities, need not, and should not, contain any clauses restricting free publication of the results. Although it has sometimes been irksome, the refusal of many universities to accept theses that cannot be published is a step in the right direction.

The future historian of science will certainly characterize the first half of the twentieth century as an age of unsurpassed progress of discovery in physics. He will also note the *crescendo* in the elaboration of physical techniques and the decisive part they played in the dramatic developments of the sister sciences. Examples are the commonplaces of our scientific practice and could be culled from almost any active region of investigation. That almost self-evident fact is well illustrated by the record of the Medallists of 1946 and I invite you to reflect on the extent to which their distinguished experimental contributions have been rendered possible by a quick appreciation of the potentialities of new physical methods. The thermionic valve, the photoelectric cell, high-vacuum technique, high-pressure technique, production and management of very low and very high temperatures, X-rays, and the use of isotopic and radioactive tracers, are but a few of the tools which modern physics has placed at our disposal.

The vastness of the subject is very significant and even if, as is necessary, I confine myself to organic chemistry, only a small part of it can be mentioned.

The forty years of my own experience have seen a revolution in the methods of experiment and unquestionably the great waves of advance are clearly identified with the introduction of new techniques. The improvement of balances and the pioneering work of Pregl brought in microanalysis and, following in its wake, microchemical manipulation. It is safe to say that this has increased the output of a given laboratory man-power by at least 100 % because of the saving of time and energy expended previously on pure routine. But even more important is the fact that microchemistry has made possible the successful attack of problems, especially in the field of biochemistry, which could not even be attempted thirty years ago.

Many of the more spectacular researches concerned substances of high biological activity and a vital part was played by the co-operation of botanists, zoologists, physiologists and bacteriologists. But equally necessary was the help of physicists in the provision of methods of investigation of 1 or 2 mg. of material.

Among the more valuable of these new resources are ultra-violet and infra-red spectroscopy and X-ray crystal analysis. The triumphs of the latter are well known and I will only add that the last details of the constitution of penicillin were revealed by the X-rays in the hands of Crowfoot and Rogers at Oxford and Bunn and Turner-Jones at Northwich. The laborious Fourier analysis which the complete mapping of electronic densities still demands, will soon be carried out by machines and it is not at all improbable that molecular structures will eventually be ascertained with ease, and almost by inspection. That will not close the organic chemical and biochemical laboratories but, on the contrary, will give impetus to their work in many fascinating directions.

Ultra-violet spectroscopy, once the concern of specialists, is now practised universally; for many purposes, however, the study of infra-red absorption promises even greater usefulness.

Although subject to constitutive influences, the bands in the infra-red are far less so than those in the ultra-violet and the method provides a kind of elementary analysis of the simpler groups contained in the molecule. It has been used *inter alia* to follow the course of polymerization, for the analysis and characterization of hydrocarbons, such as the isomeric octanes or butanes, and in the everyday control of industrial processes.

We were very impressed by, and grateful for, a recent demonstration of the power of infra-red spectroscopic analysis. A crucial test was devised in order to establish a detail of the constitution of strychnine and the outcome depended on the unequivocal identification of a degradation product obtained in very small quantity. Our own work indicated that it was carbazole mixed with one of the four C-methylcarbazoles, and probably with 3-methylcarbazole. But we could not be quite certain.

Mr Pausacker made the four methylcarbazoles, of which one was new, and Mr Richards kindly studied their infra-red spectra. They were characteristic and

differed also from that of carbazole. Using only 1.5 mg. and a novel technique, Richards showed conclusively that the specimen was essentially carbazole containing about 10% of 3-methylcarbazole. The probable course of events recalls the stages through which mountains have been said to pass—an inaccessible peak, an interesting course for experts, an easy day for a lady.

In many directions there have been notable advances in the processes of purification and analysis, but I will merely mention in passing the so-called molecular still, the ultra-centrifuge, polarography and electrophoresis.

I would however like to draw your attention to a recent series of researches which foreshadow a leap forward in our knowledge of the proteins, again because of the introduction of a new technique. In doing this I hope to make some amends for having recently bemoaned in another place the relatively small contribution of British scientists to protein research. The equipment for those who venture to follow the pioneers is not elaborate. I gather that the chief requirements are a lead tray, an earthenware drain pipe and a sheet of paper.

Although the use of animal charcoal for the removal of coloured impurities from solutions has a respectable antiquity and the separation of dyes in solution on filter paper has long been employed as a method of analysis, modern chromatography was introduced by Tswett forty years ago. He showed that coloured substances are selectively adsorbed from suitable solutions and that distinct bands are formed in a vertical column when the solution of a mixture is poured in at the top and allowed to fall through the adsorbant. In this way Tswett showed that leaf-green chlorophyll consists of two substances, later investigated by Willstätter and Stoll.

The many developments have included various devices for applying the method to colourless substances. A coloured group may be added to the molecule, the fluorescence of the bands may be observed instead of the colour, the adsorbent may be pre-coated with a fluorescent substance (Brockmann), or the column may be streaked with a reagent to produce a visible effect.

Chromatography is now a standard laboratory procedure and in this country Sir Ian Heilbron was the first to perceive its advantages.

Another well-known method of separation of substances depends on their partition between immiscible or partially immiscible solvents and an apparatus for carrying out a large number of successive partitions has been devised by L. C. Craig at the Rockefeller Institute of Medical Research.

A still more ingenious idea is that of the partition chromatography which A. J. P. Martin and R. L. M. Synge (1941) worked out in the laboratories of the Wool Industries Research Association. It makes use of a Tswett column but is based on the principles of partition rather than on that of adsorption. This distinction is evidently valid in reference to the phases concerned but it is not so certain that the two processes are not basically similar at the molecular level. As one example particles of silica gel can be impregnated with a buffer solution on the alkaline side and placed in a column through which the substances to be separated, dissolved in a suitable immiscible solvent, are passed. The effect is obviously that of a large number

of successive extractions and bands analogous to those of a chromatogram are produced; the order of the bands from top to bottom will be one of decreasing acidity of the components of the mixture. *Ceteris paribus* the most acid constituent will be found in the top layer. This technique has been found to be well adapted for the separation of the penicillins on the laboratory scale. In 1944, Martin in collaboration with R. Consden and A. H. Gordon made a further step forward by the use of water-saturated cellulose as the stationary phase and a mobile phase consisting of a solvent such as phenol or collidine, partially miscible with water. Gordon, Martin and Synge had already shown in 1943 that strips of filter paper could be used to separate amino-acids and the later work is an extension of this observation for the same purpose. The development may be one-dimensional or, preferably, two-dimensional in which procedure the first solvent is removed by drying and a second solvent is allowed to ascend the paper at right angles to the direction taken by the first. A drop of protein hydrolysate suffices and its constituent amino-acids become segregated in definite areas the position of which is dependent on the nature of the amino-acid and the solvents used. The well-known colour reaction with ninhydrin is used to show up the spots. Thus a rapid qualitative analysis of protein *bausteine* is achieved and moreover the presence of a new amino-acid will be indicated and a rough idea of its constitution will perhaps be obtained. Furthermore the simpler peptides are separated and by subsequent hydrolysis and re-partition their amino-acids can be recognized.

These researches will I believe be recognized as the greatest contribution to the study of the structure of the proteins made since the classical work of Emil Fischer.

How the method can be used is well shown by an outstanding investigation of the molecular structure of gramicidin-S by Consden, Gordon, Martin and Synge (1946). A partial hydrolysate was fractionated on two-dimensional paper chromatograms. The location of dipeptides and tripeptides having been determined these were taken from a duplicate paper, hydrolyzed, both before and after deamination, and the amino-acids identified by means of further chromatograms. The dipeptides so recognized were synthesized and their behaviour on partition paper chromatography was found to be identical with that of the respective constituents of the partial hydrolysate. The method of ionophoresis was also used and the findings were consistent. From the chain *ABCDE*, *AB*, *BC*, *CD*, *DE*, *ABC* and *DEA* were obtained and identified.

Hence, not only is the order of the five amino-acids established but it is rendered very probable that the substance is a cyclic polypeptide. The crystallographic results of Crowfoot and Schmidt are compatible with the hypothesis that the ring contains ten amino-acid groups.

At the Liverpool meeting of the British Association for the Advancement of Science (Section B, 1923) I mooted the idea that many high molecular weight substances of repeating pattern type should be regarded as mammoth rings, basing this speculation mainly on the absence of end-groups required on the open-chain hypothesis. A cyclic decapeptide would include a ring of thirty members.

I will now refer to a subject pursued in my own laboratory in collaboration with biologists, namely Dr C. E. Coulthard of the Research Department of Boots Pure Drug Company, Limited and Dr J. Ungar of Glaxo Laboratories, Limited.

The tubercle bacilli are characterized by the possession of a waxy envelope which has often been considered to confer some degree of immunity against the attack of chemotherapeutic agents. Consequently it has been sought to endow the latter with fat-soluble groups in the hope of penetrating the supposed protective covering. Actually it may be doubted whether this scheme, which has brought little success, is based on a sound conception, for it may be argued that all that could be achieved would be the establishment of a reservoir of the agent in the lipins. On these lines it would seem necessary to link the fatty part of the molecule to the water-soluble part, which it is hoped will attack the organism, by a readily hydrolyzable linkage. Several variations of this theme can be envisaged. Be this as it may, it is obvious that the chemical nature of the lipins of the bacteria deserve close attention and the first chemist to attack the problem, and with important results, was R. J. Anderson (1929 and later). The fatty acids obtained by hydrolysis of the waxes from the bacterial bodies were fractionated and one of them, tuberculostearic acid, was found by Spielman, a colleague of Anderson's to be 10-methylstearic acid.

Important constituents of the mixture were acids of the formulae $C_{26}H_{52}O_2$ and $C_{30}H_{60}O_2$; the former termed *phthioic acid* has been the more closely studied. Anderson was of the opinion that it was a branched-chain acid similar in constitution to tuberculostearic acid but the evidence garnered by him and his collaborators and by Wagner-Jauregg, was insufficient to establish the details.

E. Stenhagen and S. Stållberg then studied the behaviour of phthioic acid in monomolecular films and also the X-ray reflexions from barium phthioate. They came to the conclusion that the acid is ethyldecyldodecylacetic acid, or something very similar, but the synthesis of this substance by N. Polgar showed that this was an error probably due to the unusual degree of tilt of the molecules.

I will not burden you with the organic chemical details of Polgar's further work but combined analytic and synthetic attack of the degradation products made it very probable that phthioic acid is 3:13:19-trimethyltricosanoic acid, a straight chain of twenty-three carbon atoms with three methyl branches. Phthioic acid is feebly optically active but the optically inactive, synthetic 3:13:19-trimethyltricosanoic acid closely resembles phthioic acid in respect of its physical properties, including the behaviour of monomolecular films on water, and in the melting-points of its derivatives. We thus returned to the original general hypothesis of Anderson.

It has been known for some years that phthioic acid possesses toxic properties (F. Sabin of the Rockefeller Institute for Medical Research, New York, and others) and that it produces lesions when suitably injected into experimental animals, for example, the guinea-pig. But the observations of Coulthard and Ungar are new in that they have been able to reproduce, by a single intraperitoneal injection of synthetic acids of known constitution, a pathological picture which is almost

identical with that of tuberculosis, in respect of the particular manifestations observed. There is no doubt whatever of the reality of the phenomenon and it is highly significant.

The study of a range of synthetic branched long-chain fatty acids from this point of view is in its infancy but the following results can be cited.

The acids have been synthesized by N. Polgar, partly with the collaboration of S. David and E. Seijo.

3:12:15-Trimethyldocosanoic acid is even more active than phthioic acid, or synthetical 3:13:19-tricosanoic acid, which are equal within the limits of the method. On the other hand 2:13:17:21-tetramethyldocosanoic acid is inactive.

13:17:31-Tricosanoic acid is inactive, and so is 2:13-dimethylpentacosanoic acid.

13:16-Tricosanoic acid is active but it was suspected that the specimen contained a 3-methyl-substituted impurity. A purified specimen exhibited greatly diminished activity.

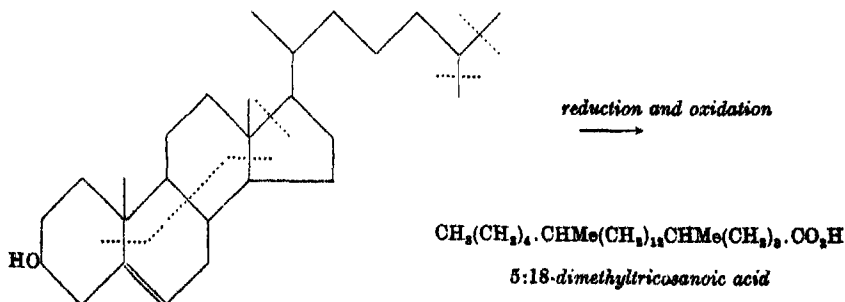
4:13:16-Tricosanoic acid is very active and though here again the presence of some 3-methyl substituent is not-excluded the activity is such that it can hardly be due to an impurity.

3:13:19- $\Delta^{13:19}$ -Tricosadienic acid is active but less so than the related saturated substance. It is probable that the specimen contains several geometrical isomerides. The syntheses are very laborious and the biological tests are prolonged, so that progress is necessarily slow.

At present it looks as if a methyl substituent in the 3- or 4-position is necessary.

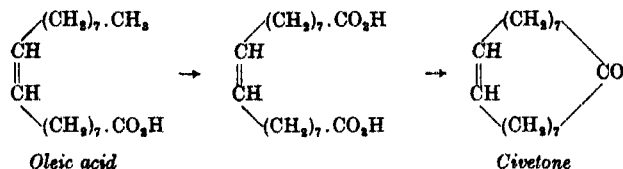
The biological property is evidently highly constitutive but it is too early to attempt an identification of all the requisite structural features. A working hypothesis is that the methyl groups block β -oxidation and some relation to physiologically active unsaturated substances may well be brought to light in the future.

An extremely interesting discovery, quite unrelated to this work in its origin, has been announced by R. P. Cook from the Biochemistry Laboratory, Cambridge University. He has obtained an acid, or a mixture of acids, $C_{28}H_{50}O_2$ by feeding cholesterol to rats. This is very suggestive of an extraordinary process of unwinding of the tetracyclic nucleus of the sterinoid by breaks at the points where the rings are fused, and also at some peripheral point, and in the side-chain. For example, one possible degradation is illustrated below:

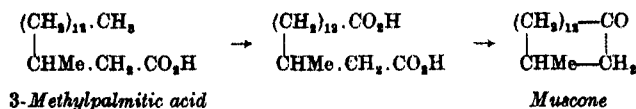


We are unable to equate the constitution of any substance that could be obtained in this way with that of phthioic acid but such a direct relation was hardly to be anticipated. If Cook's acid is really derived from the cholesterol molecule by some transformation, it must be a branched-chain acid and the determination of its structure is a most urgent problem, the solution of which must surely shed some light on an aspect of the biochemistry of phthioic acid.

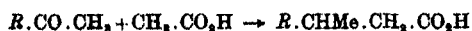
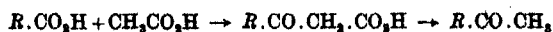
The temptation to carry speculation a little further cannot be resisted. In 1926 Professor L. Ruzicka, *For. Mem. R. S.*, elucidated the constitution of civetone, the odoriferous principle of the civet cat, and made the dramatic discovery of the existence of large carbocyclic rings in nature. He also noted the structural relation of civetone with oleic acid. Since the ketone has one less carbon atom than the acid and the latter is widely distributed, the degradation of oleic acid to civetone is more probable than the reverse synthesis. It may be suggested that oleic acid suffers ω -oxidation, a biochemical process to which Verkade has paid attention, and that this is followed by a familiar ketonization.



But analogy then leads us to assume a similar mechanism for muscone from the musk-rat, which was also studied by Ruzicka. We find that its progenitor should be a 3-methylpalmitic acid.



The occurrence of the 3-methyl-substituent is interesting in relation to phthioic acid. Further Professor Hans T. Clarke and his collaborators at Columbia University, New York, have shown by the use of C^{13} that the fatty acids are produced in the organism from acetic acid only. In parenthesis, this is a remarkable experimental justification of J. N. Collie's speculations on the role of keto-methylene chains in bio-synthesis. An additional molecule of acetic acid could be used to introduce methyl substituents by the mechanism.



3:13:19-Tricosanoic acid is not a possible product of this scheme of bio-synthesis. It would therefore not be surprising to find that phthioic acid, as at present known, is a mixture of a trimethyldocosanoic acid and a trimethyltetracosanoic acid. On this hypothesis the chain should in any case be even-numbered and the methyl groups can only be attached to the odd-numbered carbon atoms. The constitution

proposed for tuberculostearic acid conforms to the first condition, but not to the second. It could, however, be 9-methylstearic acid, if oxidation occurs at carbon atoms 9 and 10 and is accompanied by a pinacol-pinacolone migration. Alternative views to that already mentioned involve the intervention of molecules of propionic acid or formaldehyde (or an equivalent) in order to provide the methyl substituents. These, however, fix the methyls on even-numbered carbon atoms and, though tuberculostearic acid then falls into line, they are at variance with our own deductions in regard to the constitution of phthioic acid.

Following the clue afforded by chaulmoogric and hydnocarpic acids in the treatment of leprosy, Roger Adams, Davy Medallist of 1945, prepared a range of substituted fatty acids, some of which had considerable action on *B. leprae* (or possibly an analogous organism) *in vitro*. The irritating action of these substances precluded their use in practice.

Our first efforts in the field of tubercle fatty acids had a similar objective, but we have now abandoned the idea of a frontal attack on the organism in favour of an attempt to alleviate the symptoms of the disease. If this can be achieved by an immunity method it is probable that the body resistance will be strengthened. That investigation has not gone beyond the planning stage but we are glad to know that it will be in the capable hands of Professor M. Stacey at Birmingham University.

There are indications that the pathological role of abnormal lipins may not be confined to tuberculosis. Thus Novak and Grey (1938) found tuberculous tissue, with lesions, associated with granulosa cell tumours, and suggested that these effects were due to lipins produced by the malignant growths. These observations have very recently been confirmed and extended in America.

The theory of the Raman effect in crystals, in particular rock-salt

By M. BORN, F.R.S. AND MARY BRADBURN

(Received 5 October 1945)

The Raman effect in crystals is treated in this paper with the help of Placzek's approximation. It consists of contributions of different orders with respect to the amplitudes of the vibrations; the first-order effect is a line spectrum depending only on the vibrations of infinite wavelength, the second-order effect is a continuous spectrum depending on combination frequencies of all pairs of branches of the lattice vibrations, each pair taken for the same wave vector. In highly symmetrical crystals like rock-salt the first-order effect is zero. The second-order effect can be calculated for rock-salt with the help of the tables of the lattice frequencies published by Kellermann. It consists of thirty-six peaks, each belonging to a combination frequency. The superposition of these allows us to determine without any arbitrary assumption about the coupling constants, the frequency of the observable maxima in fair agreement with Krishnan's measurements. By adapting three coupling constants one can also determine the relative intensities of the most prominent peaks and obtain a curve which in its main features agrees with the observed one.

The results show that lattice dynamics can account quantitatively for the Raman effect in crystals and that Raman's attacks against the theory are unfounded.

INTRODUCTION

In the same year, 1928, as Raman discovered the existence of displaced lines in the light scattered by liquids, the Russian physicists, Landsberg and Mandelstam, found a similar phenomenon in crystals (quartz and calcite). The Raman effect provided a powerful tool for investigating the dynamical properties of molecules, and a detailed theory of molecular scattering in liquids and gases has been developed. Crystals have been comparatively neglected, and no systematic theory of their Raman effect seems to exist. The reason for this is easily understood. In those crystals which can be considered as composed of molecules the effect is roughly identical with that in the liquid state, simplified to some extent by the fixed orientation, but complicated by the coupling of the molecules; and as the main interest is directed towards using the Raman spectrum as a help to the chemist, the embedding of the molecule into a crystal appears as an unnecessary complication and nuisance. In atomic and ionic crystals, however, where the whole lattice has to be regarded as a single gigantic molecule, the intensity of the Raman effect is very often extremely small; for instance, when certain symmetry elements exist which produce a cancelling of terms of first order with respect to the vibrational amplitudes.

With the help of very long exposures, however, the second-order Raman effect can be observed. The first example of this kind was obtained by Rasetti (1931) and discussed by Fermi & Rasetti (1931). They found that the Raman effect in rock-salt crystals consists of a continuous background with peaks superimposed, and they developed the theory from the standpoint of ordinary lattice dynamics to a stage

where the continuous character of the spectrum could be understood; they abstained, however, from any attempt at a quantitative theory as their formulae were too involved.

We have now taken up this problem and developed the theory for rock-salt in some detail because Rasetti's observations have been used by Raman and his school as material for discrediting lattice dynamics and for proving their own strange theory of lattice vibrations. This theory contends that the vibrational spectrum is not quasi-continuous as correct classical or quantum mechanics implies, but consists of a small number of lines. All kinds of phenomena have been used by Raman for demonstrating this statement: specific heat, diffuse X-ray scattering, optical effects. The temperature dependence of the specific heat is, of course, not sensitive enough to supply a sharp criterion; it is, for example, well known that Lindemann's formula, which uses only two frequencies, can hardly be distinguished from Debye's formula, which uses a continuous spectrum. In the case of diffuse X-ray scattering the situation is now even less favourable for Raman's ideas; for it turns out that the diffuse thermal X-ray spots, when correctly interpreted, are a direct image of the continuous vibrational spectrum, which could in principle be derived from measurements of the intensity distribution of the scattering. There remains the optical effects, which, according to Raman, show directly the existence of only a small number of characteristic frequencies in crystals. Raman's pupil Krishnan (1943) has written a paper on the Raman effect in rock-salt, discussing and reinterpreting the observations of Fermi & Rasetti. He contends that their result has to be regarded as a line spectrum (of 9 Stokes and 9 anti-Stokes lines). Now the photograph of the spectrum may possibly be interpreted as a system of faint lines, though the continuous background seems obvious; Fermi & Rasetti have, however, published above the photograph the microphotometric curve of the intensity distribution which shows without the slightest doubt that the Italian authors are right in describing it as a continuum with small peaks. Krishnan says that the 9 lines, which he counts rather arbitrarily on each side of the incident line, are exactly what Raman's theory predicts; but he makes no attempt to calculate the position of the lines from his theory nor to discuss the intensities.

Recently, Krishnan (1945*a*) has repeated the observations and obtained a spectrum and microphotometer curve which is a confirmation of Rasetti's work and an improvement on it in so far as the little peaks are much sharper. Krishnan declares again that they constitute a line spectrum, but he counts only 6 strong and 2 weak lines, some of which may be double. So it seems that he has given up the claim that these peaks confirm the existence of 9 lines as predicted by Raman's theory.

Under these circumstances it seemed to us desirable to develop the correct lattice theory of the Raman effect for crystals and apply it quantitatively to rock-salt. This substance is particularly suited for this purpose, since Kellermann (1940) has published the theory of the lattice vibrations of rock-salt containing tables of the coefficients of the equations of motion and of the six frequency branches for a fairly narrow distribution of wave vectors.

We give here in §§ 1–5 the general theory for any crystal. The main result is that the second-order Raman effect is a continuous spectrum involving combinations $\omega_j \pm \omega_{j'}$ of two frequencies in the same or different branches, but both belonging to the same wave vector. The peaks on this continuous background can be understood by a principle already used by Born & Blackman (1933) for explaining the fine structure of the residual rays† (*Reststrahlen*), namely, that the density of the frequencies of a lattice has a maximum near the centre of the reciprocal cell. The Raman spectrum consists, in fact, of a number of steep and narrow continuous peaks which are superimposed in such a way that the observed aspect is obtained.

If this general theory is applied to rock-salt a satisfactory representation of the extent of the continuous background and the position of the main peaks is obtained; no arbitrary constants are introduced in the determination of these, as Kellermann's results depend only on the measured lattice constant and compressibility of rock-salt. In order to explain the intensities the coupling constants between the ions of the lattice have to be adapted.

Krishnan has published some other second-order Raman effects, e.g. in diamond (1945*b*), obtained by extremely long exposures, which look superficially like faint lines. One can expect that these will also be explained quantitatively by a correct application of lattice dynamics.

1. THE GENERAL THEORY OF THE RAMAN EFFECT

A perfectly rigorous theory of crystal optics should proceed in this way: The quantized radiation field should be coupled to a set of nuclei and the corresponding number of electrons in an almost periodic array; the whole has to be treated as one quantum-mechanical system. This would, however, be extremely complicated. For the purpose of calculating the scattering the problem can be simplified by assuming the radiation enclosed in a very large box and the crystal small compared with the box, yet, on the other hand, large compared with the lattice cell. Then the whole crystal can be considered as one gigantic molecule and Dirac's theory of the interaction between particles and radiation can be applied. The result can best be described by first considering the scattering classically.

Assume that the incident light is elliptically polarized,

$$\mathbf{E} = \Re(\mathbf{A}e^{-i\omega t}) = \frac{1}{2}(\mathbf{A}e^{-i\omega t} + \mathbf{A}^*e^{i\omega t}); \quad (1.1)$$

it produces an electric moment \mathbf{M} with components

$$M_\rho = \sum_\sigma \alpha_{\rho\sigma} A_\sigma \quad (\rho, \sigma = 1, 2, 3), \quad (1.2)$$

† This theory was based on the assumption of non-harmonic terms in the potential energy. One can, however, explain the observations more simply without this assumption, using only harmonic vibrations, by taking into account the deformability of the particles. This will be shown in a separate publication.

where $\alpha_{\rho\sigma}$ is the polarizability tensor. Let \mathbf{q} be a unit vector normal to the direction of observation. The intensity of the light scattered in this direction is, apart from constant factors,

$$I = \omega^4 |\mathbf{M} \cdot \mathbf{q}|^2 = \omega^4 \sum_{\rho\sigma} \sum_{\mu\nu} \alpha_{\rho\sigma} \alpha_{\mu\nu}^* A_\sigma A_\nu^* q_\rho q_\mu. \quad (1.3)$$

In quantum theory the $\alpha_{\rho\sigma}$ can be calculated in terms of the frequency ω of the incident light, the energy levels of the scattering system and the matrix elements of its natural or unperturbed electric moment. The light scattered for the transition of the system from a quantum state n to another n' is obtained by replacing $\alpha_{\rho\sigma}$ in formula (1.3) by the matrix element $[\alpha_{\rho\sigma}]_{nn'}$ and ω^4 by $(\omega + \omega_{nn'})^4$.

Practically only vibrational transitions are observable, producing the Rayleigh and Raman effect. In this case a simplification due to Placzek (1934) can be used. The polarizability α can be calculated as if the nuclei were in fixed positions (neglecting their kinetic energy) so that it is a function of the nuclear configuration chosen, $\alpha_{\rho\sigma}(X)$. Now if the nuclei are performing small vibrations (U) about equilibrium positions (X_0) so that $X = X_0 + U$, the vibrational states may be characterized by the quantum number v and the matrix elements of $\alpha_{\rho\sigma}(U)$ with respect to v , $[\alpha_{\rho\sigma}]_{vv'}$, can be formed. Then (1.3) still holds for these coefficients. In this approximation $(\omega + \omega_{nn'})^4$ can be neglected, as it is practically the same for all Raman lines of a given incident light:

$$I_{vv'} = \sum_{\rho\sigma} \sum_{\mu\nu} [\alpha_{\rho\sigma}]_{vv'} [\alpha_{\mu\nu}^*]_{vv'} A_\sigma A_\nu^* q_\rho q_\mu. \quad (1.4)$$

We shall specialize the general formulae for natural incident light and observations without an analyser. In this case the average over the azimuth of the polarization of the incident and scattered light has to be taken. If θ_ρ ($\rho = 1, 2, 3$) are the angles between the incident beam and the co-ordinate axes and ϕ_ρ ($\rho = 1, 2, 3$) those between the observed beam and these axes, an elementary consideration leads to

$$\left. \begin{aligned} \overline{A_\rho A_\sigma^*} &= -I_0 \cos \theta_\rho \cos \theta_\sigma \quad (\rho \neq \sigma), & \overline{A_\rho A_\rho^*} &= I_0 \sin^2 \theta_\rho, \\ \overline{q_\rho q_\sigma} &= -\frac{1}{2} \cos \phi_\rho \cos \phi_\sigma \quad (\rho \neq \sigma), & \overline{q_\rho q_\rho} &= \frac{1}{2} \sin^2 \phi_\rho, \end{aligned} \right\} \quad (1.5)$$

where $I_0 = \frac{1}{2} |\mathbf{A}|^2 = E^2$ is the intensity of the incident light.

The problem is now reduced to the calculation of the quantities

$$[i_{\rho\sigma, \mu\nu}]_{vv'} = [\alpha_{\rho\sigma}]_{vv'} [\alpha_{\mu\nu}^*]_{vv'} \quad (1.6)$$

for all vibrational transitions $v \rightarrow v'$; but very many of these belong to the same frequency, particularly if the vibrations are considered to be harmonic (as we shall do in our case). Then the weighted mean of the quantity (1.6) has to be formed.

Let ϵ_v be the energy in the state v , then by multiplying by the Boltzmann factor the thermal average

$$\langle i_{\rho\sigma, \mu\nu} \rangle_{av} = \frac{\sum_v [i_{\rho\sigma, \mu\nu}]_{vv'} e^{-\epsilon_v/kT}}{\sum_v e^{-\epsilon_v/kT}}, \quad (1.7)$$

where the summation is taken over the initial state v . The result can be ordered with respect to the oscillator frequencies separating the state v' from the state v .

2. EXPANSION OF THE POLARIZABILITY

We describe the equilibrium of the crystal in the usual way with the help of the base index k and the cell index $l(l_1, l_2, l_3)$. If n is the number of particles in the cell, k assumes n values $0, 1, 2, \dots, n-1$; if N is the number of cells the domain of l consists of N points. We denote by $\mathbf{u} \begin{pmatrix} l \\ k \end{pmatrix}$ a small displacement of the nucleus (l, k) and write its three rectangular components $u_\mu \begin{pmatrix} l \\ k \end{pmatrix}$ ($\mu = 1, 2, 3$). The whole domain of values μ, l, k is $3nN$.

The polarizability can then be expanded as a power series with respect to the u ,

$$\alpha_{\rho\sigma} = \alpha_{\rho\sigma}^0 + \alpha_{\rho\sigma}^{(1)} + \alpha_{\rho\sigma}^{(2)} + \dots, \quad (2.1)$$

where

$$\left. \begin{aligned} \alpha_{\rho\sigma}^0 &= \text{constant}, \\ \alpha_{\rho\sigma}^{(1)} &= \sum_{\mu} \sum_{kl} \alpha_{\rho\sigma, \mu} \begin{pmatrix} l \\ k \end{pmatrix} u_{\mu} \begin{pmatrix} l \\ k \end{pmatrix}, \\ \alpha_{\rho\sigma}^{(2)} &= \sum_{\mu\nu} \sum_{kl} \sum_{k'l'} \alpha_{\rho\sigma, \mu\nu} \begin{pmatrix} ll' \\ kk' \end{pmatrix} u_{\mu} \begin{pmatrix} l \\ k \end{pmatrix} u_{\nu} \begin{pmatrix} l' \\ k' \end{pmatrix}. \end{aligned} \right\} \quad (2.2)$$

All these quantities are symmetric in ρ, σ but not in the other indices.

The coefficients satisfy a number of identities. The first set of these are consequences of the periodicity of the lattice. Any quantity which describes an equilibrium property and depends on only one single particle (l, k) must be the same for all cells and hence independent of l ; in particular

$$\alpha_{\rho\sigma, \mu} \begin{pmatrix} l \\ k \end{pmatrix} \text{ independent of } l, = \alpha_{\rho\sigma, \mu}(k). \quad (2.3)$$

Further, any equilibrium quantity depending on two points (l, k) and (l', k') is invariant if l and l' are subjected to the same cell displacement and hence depends only on the difference $l - l' (l_1 - l'_1, l_2 - l'_2, l_3 - l'_3)$; in particular

$$\alpha_{\rho\sigma, \mu\nu} \begin{pmatrix} ll' \\ kk' \end{pmatrix} = \alpha_{\rho\sigma, \mu\nu} \begin{pmatrix} l-l' \\ kk' \end{pmatrix}. \quad (2.4)$$

A second set of identities is obtained from the consideration that any translation of the crystal as a whole does not change its electric moment. Hence if all $\mathbf{u} \begin{pmatrix} l \\ k \end{pmatrix}$ are taken equal (independent of l and k) $\alpha_{\rho\sigma}^{(1)}, \alpha_{\rho\sigma}^{(2)}$ must vanish. Therefore

$$\sum_k \alpha_{\rho\sigma, \mu}(k) = 0, \quad \sum_l \sum_{kk'} \alpha_{\rho\sigma, \mu\nu} \begin{pmatrix} l \\ kk' \end{pmatrix} = 0. \quad (2.5)$$

3. NORMAL CO-ORDINATES

As we have already said, we shall consider the vibrations to be harmonic. Let $\Phi(X)$ be the total electronic energy for fixed nuclei which can be expanded with respect to the displacements U in the neighbourhood of the equilibrium position X_0 ,

$$\Phi(X) = \Phi(X_0 + U) = \Phi_0 + \Phi_1 + \Phi_2 + \dots, \quad (3.1)$$

where Φ_r is of degree r in U . Φ_1 vanishes in equilibrium. Φ_2 can be written

$$\Phi_2 = \frac{1}{2} \sum_{\mu\nu} \sum_{kl} \sum_{k'l'} \Phi_{\mu\nu} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} u_\mu \begin{pmatrix} l \\ k \end{pmatrix} u_\nu \begin{pmatrix} l' \\ k' \end{pmatrix}, \quad (3.2)$$

where the coefficients are the equilibrium values of the second derivatives,

$$\frac{\partial^2 \Phi}{\partial u_\mu \begin{pmatrix} l \\ k \end{pmatrix} \partial u_\nu \begin{pmatrix} l' \\ k' \end{pmatrix}} = \Phi_{\mu\nu} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} = \sqrt{(m_k m_{k'})} \mathcal{D}_{\mu\nu} \begin{pmatrix} l-l' \\ kk' \end{pmatrix}, \quad (3.3)$$

which depend on two points and therefore on the difference $l-l'$ only. \mathcal{D} considered as a function of the two sets of indices μ, l, k and ν, l', k' is called the dynamical matrix.

Now the quantum theory of electronic systems (Born & Oppenheimer 1927) shows that the total electronic energy can be considered to be the potential energy for the motion of the nuclei. Their kinetic energy is

$$\mathcal{T} = \frac{1}{2} \sum_{kl} m_k \dot{u} \begin{pmatrix} l \\ k \end{pmatrix}^2. \quad (3.4)$$

Now introduce normal co-ordinates with the help of which Φ_2 and \mathcal{T} are transformed into the sums of squares. We give only the results of the well-known method. Let $\xi(J)$ be complex normal co-ordinates where the index J assumes the same number of values $3nN$ as the domains of μ, k, l together. Then one has

$$\left. \begin{aligned} \sqrt{(m_k)} u_\mu \begin{pmatrix} l \\ k \end{pmatrix} &= \sum_J e_\mu \begin{pmatrix} l \\ k \end{pmatrix} J \xi(J), \\ \xi(J) &= \sum_{\mu, l, k} e_\mu^* \begin{pmatrix} l \\ k \end{pmatrix} J \sqrt{(m_k)} u_\mu \begin{pmatrix} l \\ k \end{pmatrix}. \end{aligned} \right\} \quad (3.5)$$

The coefficients are the solutions of the linear equations for free vibrations (waves). In consequence of the periodicity of the lattice these can be split into N sets of $3n$ equations,

$$\omega^2 \begin{pmatrix} Q \\ j \end{pmatrix} e_\mu \begin{pmatrix} k \\ j \end{pmatrix} Q = \sum_\nu \sum_{k'} \mathcal{D}_{\mu\nu}(kk' | Q) e_\nu \begin{pmatrix} k' \\ j \end{pmatrix} Q, \quad (3.6)$$

$$\text{where} \quad \mathcal{D}_{\mu\nu}(kk' | Q) = \sum_l \mathcal{D}_{\mu\nu} \begin{pmatrix} l \\ kk' \end{pmatrix} e^{-2\pi i l \cdot Q}, \quad (l, Q) = l_1 Q_1 + l_2 Q_2 + l_3 Q_3. \quad (3.7)$$

$$\text{Since obviously} \quad \mathcal{D}_{\mu\nu}(kk' | -Q) = \mathcal{D}_{\mu\nu}^*(kk' | Q), \quad (3.8)$$

$$\text{it follows that} \quad e_\mu \begin{pmatrix} k \\ j \end{pmatrix} -Q = e_\mu^* \begin{pmatrix} k \\ j \end{pmatrix} Q. \quad (3.9)$$

The index J has been split into two indices $\begin{pmatrix} Q \\ j \end{pmatrix}$, where $j = 1, 2, \dots, 3n$ represents the different branches of the spectrum and $Q(Q_1, Q_2, Q_3)$ the different waves of each branch. The full set of different values of Q is represented by the points which are obtained on dividing any unit cell in reciprocal space into $N = L^3$ sub-cells each of side $1/L$.

The coefficients in (3.5) are the elements of a unitary matrix and can be written

$$e_{\mu}\left(\begin{matrix} l \\ k \end{matrix} \middle| J\right) = e_{\mu}\left(\begin{matrix} l \\ k \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) = e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) e^{2\pi i t(l, Q)}, \quad (3.10)$$

where the $e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right)$ must satisfy the identities

$$\sum_{\mu, k} e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) e_{\mu}^*\left(\begin{matrix} k \\ j' \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) = \delta(jj'), \quad \sum_j e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) e_{\nu}^*\left(\begin{matrix} k' \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) = \delta(\mu\nu) \delta(kk'). \quad (3.11)$$

Since the $u\left(\begin{matrix} l \\ k \end{matrix}\right)$ are real, one has

$$\xi\left(\begin{matrix} -Q \\ j \end{matrix}\right) = \xi^*\left(\begin{matrix} Q \\ j \end{matrix}\right). \quad (3.12)$$

Now each complex ξ represents two real normal co-ordinates, but on account of (3.12) the number of these can be reduced to the proper value again by restricting the Q values to half the points in the unit cell of reciprocal space.

4. EXPRESSION OF POLARIZABILITY IN TERMS OF NORMAL CO-ORDINATES

Substituting (3.5) in (2.2) one obtains

$$\begin{aligned} \alpha_{\rho\sigma}^{(1)} &= \sum_{\mu} \sum_{lk} \alpha_{\rho\sigma, \mu}(k) u_{\mu}\left(\begin{matrix} l \\ k \end{matrix}\right) \\ &= \sum_j \sum_{\mu} \sum_{lk} \alpha_{\rho\sigma, \mu}(k) e^{2\pi i t(l, Q)} e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} Q \\ j \end{matrix}\right) \xi\left(\begin{matrix} Q \\ j \end{matrix}\right) \frac{1}{\sqrt{m_k}}. \end{aligned} \quad (4.1)$$

Here the factor $\sum_l e^{2\pi i t(l, Q)}$ splits off; it is a δ -function vanishing except for $Q = 0$.

$$\text{Hence} \quad \alpha_{\rho\sigma}^{(1)} = \sum_j \sum_{\mu} \sum_k \alpha_{\rho\sigma, \mu}(k) e_{\mu}\left(\begin{matrix} k \\ j \end{matrix} \middle| \begin{matrix} 0 \\ j \end{matrix}\right) \xi\left(\begin{matrix} 0 \\ j \end{matrix}\right) \frac{1}{\sqrt{m_k}}. \quad (4.2)$$

The first-order Raman effect is therefore due only to vibrations with wave number zero or infinite wave-length, where each of the n simple lattices corresponding to the n points of the base moves like a rigid system. The matrix elements $\left[\xi\left(\begin{matrix} 0 \\ j \end{matrix}\right)\right]_{vv'}$ vanish except for transitions belonging to the frequencies $\pm \omega\left(\begin{matrix} 0 \\ j \end{matrix}\right)$; therefore the first-order Raman effect is a line spectrum of not more than $3n - 3$ lines (as the three acoustical branches have $\omega\left(\begin{matrix} 0 \\ j \end{matrix}\right) = 0$), each line corresponding to a normal mode of the system of n interpenetrating rigid simple lattices. It is the same as the Raman effect of a molecule of n particles, with a fixed orientation in space (translations allowed, rotations excluded) having the dynamical matrix (3.7)

$$\mathcal{D}_{\mu\nu}(kk' | 0) = \sum \mathcal{D}_{\mu\nu}\left(\begin{matrix} l \\ kk' \end{matrix}\right).$$

But (4.2) always vanishes if each lattice point is a centre of symmetry. This is the case in many simple crystals, among them rock-salt. Therefore we shall not discuss the first-order Raman effect any further.

In the same way it follows from (2.2) and (3.5) that

$$\begin{aligned}\alpha_{\rho\sigma}^{(2)} &= \sum_{\mu\nu} \sum_{lk} \sum_{l'k'} \alpha_{\rho\sigma, \mu\nu} \binom{l-l'}{kk'} u_{\mu} \binom{l}{k} u_{\nu} \binom{l'}{k'} \\ &= \sum_Q \sum_{Q'} \sum_{\mu\nu} \sum_{lk} \sum_{l'k'} \alpha_{\rho\sigma, \mu\nu} \binom{l-l'}{kk'} e^{2\pi i(l, Q)+(l', Q')} e_{\mu} \left(k \middle| \begin{smallmatrix} Q \\ j \end{smallmatrix} \right) e_{\nu} \left(k' \middle| \begin{smallmatrix} Q' \\ j' \end{smallmatrix} \right) \xi \left(\begin{smallmatrix} Q \\ j \end{smallmatrix} \right) \xi \left(\begin{smallmatrix} Q' \\ j' \end{smallmatrix} \right) \frac{1}{\sqrt{(m_k m_{k'})}}.\end{aligned}\quad (4.3)$$

Replacing $l-l'$ by l the factor $\sum_{l'} e^{2\pi i(l', Q+Q')}$ splits off; it is a δ -function of $Q+Q'$, hence in the sum (4.3) only the terms $Q' = -Q$ appear. Using (3.5) and (3.9)

$$\alpha_{\rho\sigma}^{(2)} = \sum_Q \sum_{jj'} \alpha_{\rho\sigma} \left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix} \right) \xi \left(\begin{smallmatrix} Q \\ j \end{smallmatrix} \right) \xi^* \left(\begin{smallmatrix} Q \\ j' \end{smallmatrix} \right), \quad (4.4)$$

$$\text{where} \quad \alpha_{\rho\sigma} \left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix} \right) = \sum_l \sum_{kk'} \sum_{\mu\nu} \alpha_{\rho\sigma, \mu\nu} \binom{l}{kk'} e^{2\pi i(l, Q)} e_{\mu} \left(k \middle| \begin{smallmatrix} Q \\ j \end{smallmatrix} \right) e_{\nu}^* \left(k' \middle| \begin{smallmatrix} Q \\ j' \end{smallmatrix} \right) \frac{1}{\sqrt{(m_k m_{k'})}}. \quad (4.5)$$

5. THERMAL AVERAGES AND THE INTENSITY DISTRIBUTION

Now substitute (4.4) for $\alpha_{\rho\sigma}$ in (1.6) giving

$$[i_{\rho\sigma, \mu\nu}]_{vv'} = \sum_Q \sum_{jj'} \alpha_{\rho\sigma} \left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix} \right) \left[\xi \left(\begin{smallmatrix} Q \\ j \end{smallmatrix} \right) \xi^* \left(\begin{smallmatrix} Q \\ j' \end{smallmatrix} \right) \right]_{vv'} \sum_{Q'} \sum_{j''j'''} \alpha_{\mu\nu}^* \left(\begin{smallmatrix} Q' \\ j''j''' \end{smallmatrix} \right) \left[\xi^* \left(\begin{smallmatrix} Q' \\ j'' \end{smallmatrix} \right) \xi \left(\begin{smallmatrix} Q' \\ j''' \end{smallmatrix} \right) \right]_{vv'}. \quad (5.1)$$

This expression is then to be inserted in (1.7) and the result will obviously be a sum of terms of the type

$$\left\langle \left[\xi \left(\begin{smallmatrix} Q \\ j \end{smallmatrix} \right) \xi^* \left(\begin{smallmatrix} Q \\ j' \end{smallmatrix} \right) \right]_{vv'} \left[\xi^* \left(\begin{smallmatrix} Q' \\ j'' \end{smallmatrix} \right) \xi \left(\begin{smallmatrix} Q' \\ j''' \end{smallmatrix} \right) \right]_{vv'} \right\rangle_{av}. \quad (5.2)$$

The quantum number v represents the set v_1, v_2, v_3, \dots of the quantum numbers of the single oscillators.

It is well known that the matrix elements $[\eta_J]_{vv'}$ of the amplitude η_J of one of a set of real oscillators vanish except when v_J changes by ± 1 while all other $v_{J'}$ are unchanged. The matrix elements of the square η_J^2 vanish except when v_J changes by 0, ± 2 and those of the product $\eta_J \eta_{J'}$ ($J \neq J'$), except when v_J and $v_{J'}$ change simultaneously by ± 1 .

The results can be ordered according to the frequencies

$$\omega_{vv'} = \sum_J \omega(J) (v_J - v_{J'}), \quad (5.3)$$

and are given in table 1, where $c_J = \sqrt{[\hbar/2\omega(J)]}$.

TABLE I

	0	$2\omega(J)$	$-2\omega(J)$
$\frac{\eta_J \eta_{J'}}{\eta_J^2}$	$c_J^2(2v_J + 1)$	$c_J^2 \sqrt{[(v_J + 1)(v_J + 2)]}$	$c_J^2 \sqrt{[v_J(v_J - 1)]}$
	$\omega(J) + \omega(J')$	$-\omega(J) + \omega(J')$	$\omega(J) - \omega(J')$
$\frac{\eta_J \eta_{J'}}{\eta_J^2}$	$c_J c_{J'} \sqrt{[(v_J + 1)(v_{J'} + 1)]}$	$c_J c_{J'} \sqrt{[v_J v_{J'}]}$	$c_J c_{J'} \sqrt{[v_{J'}(v_J + 1)]}$

A product of the form (5.2) with real oscillators, e.g.

$$\langle [\eta_J \eta_{J'}]_{vv'} [\eta_{J''} \eta_{J'''}]_{vv'} \rangle_{av}, \quad (5.4)$$

where the two factors refer to the same transition $v \rightarrow v'$, is obviously zero except when the first factor refers to the same two oscillators as the second, i.e. if either

$$J = J'', J' = J''' \quad \text{or} \quad J = J''', J' = J''. \quad (5.5)$$

Now consider a system of complex oscillators

$$\xi(J) = \eta + i\zeta, \quad (5.6)$$

where η is short for η_J , then the last result still applies.

Consider first $J = J' = J'' = J'''$, then

$$\langle [\xi(J) \xi^*(J)]_{vv'} [\xi(J) \xi(J)]_{vv'} \rangle_{av} = \langle [\xi \xi^*]_{vv'}^2 \rangle_{av}, \quad (5.7)$$

where we have omitted the argument J . Now consider only transitions $v \rightarrow v'$ corresponding to the Raman effect (not the Rayleigh effect, i.e. excluding $v' = v$). Then

$$\langle [\xi \xi^*]_{vv'}^2 \rangle_{av} = \langle [\eta^2]_{vv'}^2 + [\zeta^2]_{vv'}^2 + 2[\eta^2]_{vv'} [\zeta^2]_{vv'} \rangle_{av}. \quad (5.8)$$

Here the last term vanishes; for the meaning of the brackets is explicitly the following:

$$\begin{aligned} [\eta^2]_{vv'} &= [\eta^2]_{n, n \pm 2} \delta(n, n' \mp 2) \delta(m, m'), \\ [\zeta^2]_{vv'} &= [\zeta^2]_{m, m \pm 2} \delta(n, n') \delta(m, m' \mp 2), \end{aligned} \quad (5.9)$$

where n is the quantum number belonging to η , m that belonging to ζ , and $v' \neq v$. Hence the product vanishes. Since, however, the two matrix elements which appear in the other terms belong to the same transition frequency, (5.8) finally reduces to

$$\langle [\xi \xi^*]_{vv'}^2 \rangle_{av} = 2\langle [\eta^2]_{vv'}^2 \rangle_{av}. \quad (5.10)$$

In the same way, if $J = J'', J' = J''', J \neq J'$,

$$\begin{aligned} \langle [\xi(J) \xi^*(J')]_{vv'} [\xi^*(J) \xi(J')]_{vv'} \rangle_{av} &= \langle [\eta \eta']_{vv'}^2 + [\zeta \zeta']_{vv'}^2 + [\eta \zeta']_{vv'}^2 + [\eta' \zeta]_{vv'}^2 \rangle_{av} \\ &= 4\langle [\eta \eta']_{vv'}^2 \rangle_{av}, \end{aligned} \quad (5.11)$$

and in the third case $J = J''', J' = J''; J \neq J'$,

$$\begin{aligned} \langle [\xi(J) \xi^*(J')]_{vv'} [\xi(J) \xi^*(J')]_{vv'} \rangle_{av} &= \langle [\eta \eta']_{vv'}^2 + [\zeta \zeta']_{vv'}^2 - [\eta \zeta']_{vv'}^2 - [\eta' \zeta]_{vv'}^2 \rangle_{av} \\ &= 0. \end{aligned} \quad (5.12)$$

Finally, the averages for the real oscillators occurring in the formulae (5.10) and (5.11) have to be calculated. With the abbreviation

$$\beta(J) = \frac{\hbar\omega(J)}{kT}, \quad \mathcal{C}(J) = \frac{c_J^2}{1 - e^{-\beta(J)}}, \quad (5.13)$$

straightforward calculations from (1.7) and table 1 lead to the results shown in table 2.

TABLE 2

	$\langle [\eta_J^2]_{vv} \rangle_{av}$		$\langle [\eta_J \eta_{J'}]_{vv} \rangle_{av}$
$2\omega(J)$	$2\mathcal{C}(J)^2 e^{-\frac{1}{2}\beta(J)}$	$\omega(J) + \omega(J')$	$\mathcal{C}(J) \mathcal{C}(J') e^{-\beta(J) - \beta(J')}$
$-2\omega(J)$	$2\mathcal{C}(J)^2$	$-\omega(J) - \omega(J')$	$\mathcal{C}(J) \mathcal{C}(J')$
		$\omega(J) - \omega(J')$	$\mathcal{C}(J) \mathcal{C}(J') e^{-\beta(J)}$

The intensities corresponding to these six frequencies are obtained from (5.1) with the help of (5.10), (5.11) and (5.12). Writing

$$\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) = \omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \pm \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right),$$

the Stokes lines (St.) correspond to the frequency shift $-\left|\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)\right|$ and the anti-Stokes (A.St.) to $+\left|\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)\right|$. The frequencies and the intensities of single emissions are given by the following scheme:

$$\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) = \begin{cases} \omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) + \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right) \\ \omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) - \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right) \end{cases} \quad i_{\rho\sigma, \mu\nu}\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) = 2\mathcal{C}\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \mathcal{C}\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right) \left\{ \alpha_{\rho\sigma}\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) \alpha_{\mu\nu}^*\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) \right. \\ \left. + \alpha_{\rho\sigma}\left(\begin{smallmatrix} Q \\ j'j \end{smallmatrix}\right) \alpha_{\mu\nu}^*\left(\begin{smallmatrix} Q \\ j'j \end{smallmatrix}\right) \right\} \begin{cases} 1 & \text{St.} \\ \exp\left[-\beta\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) - \beta\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right)\right] & \text{A.St.} \end{cases} \quad (5.14)$$

To obtain the observed intensities one must first integrate over that part of the reciprocal space Q which belongs to a frequency ω , i.e.

$$i_{\rho\sigma, \mu\nu}(\omega) d\omega = \sum_{jj'} \iiint_{\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) < \omega + d\omega} i_{\rho\sigma, \mu\nu}\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right) dQ_1 dQ_2 dQ_3, \quad (5.15)$$

where the Stokes and anti-Stokes lines have to be taken separately. Then the total intensity for a given incident beam A and a given position of the analyser q is, from (1.4),

$$I(\omega) = \sum_{\rho\sigma, \mu\nu} i_{\rho\sigma, \mu\nu}(\omega) A_\rho A_\mu^* q_\sigma q_\nu. \quad (5.16)$$

The number of terms in the sum (5.15) over j and j' is $(3n)^2$; for instance, for a diatomic base it is 36. Hence the whole observable spectrum consists of $9n^2$ superimposed continuous bands each of which will have a maximum at a certain point. This result explains at once the general feature of the observed spectrum; that it consists of a continuous background with a considerable number of peaks.

We shall not continue the general theory any further except to indicate a few points in the evaluation of the intensity. Symmetry considerations can be used to simplify the α -factors occurring in (5.14) using the explicit expressions (4.5). This would be extremely involved if it were to be done for all waves, i.e. for all points of the reciprocal space Q . In fact, the integral (5.15) will depend very little on the whole distribution of $i_{\rho\sigma,\mu\nu}\left(\frac{Q}{jj'}\right)$ in the Q -space, but mainly on those parts where the frequencies have a maximum of density. The general problem of finding these regions of the Q -space would also be very involved (we shall solve it approximately in the special case of rock-salt). If it is solved it would suffice to apply the symmetry considerations to these special waves which would be a much simpler problem than the general one and would lead to a classification of the qualitative feature of the Raman spectra of crystals according to the symmetry of the lattice.

6. RAMAN EFFECT IN ROCK-SALT

For the rock-salt lattice the whole vibration spectrum has been calculated by Kellermann (1940) for a set of Q -values sufficiently complete to get a good idea of all the branches. His assumptions are Coulomb forces acting between all ions and repulsive forces between next neighbours; the two constants appearing from the latter can be calculated by using the experimental values of the lattice constant and the compressibility. Hence his results are independent of arbitrary assumptions, and they are checked by the calculations of the specific heat (Kellermann 1941).

First consider the factor containing the polarizability α in (5.14), assuming that the constants $\alpha_{\rho\sigma,\mu\nu}\left(\frac{l}{kk'}\right)$ in (4.5) are different from zero only for next neighbours; even under this simplified assumption seven independent constants appear. For rock-salt, a diatomic cubic crystal, k and k' , are equal to either 1 or 2, and there are six next neighbours:

l	1	2	3	4	5	6
l_1	1	0	0	-1	0	0
l_2	0	1	0	0	-1	0
l_3	0	0	1	0	0	-1

From general symmetry considerations

$$\alpha_{\rho\sigma,\mu\nu}\left(\frac{l}{kk'}\right) = \alpha_{\rho\sigma,\mu\nu}\left(\frac{-l}{k'k}\right), \quad (6.1)$$

and since any lattice point of rock-salt is a centre of symmetry

$$\alpha_{\rho\sigma,\mu\nu}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) = \alpha_{\rho\sigma,\mu\nu}\left(\begin{smallmatrix} -l \\ kk' \end{smallmatrix}\right). \quad (6.2)$$

The following equations give the seven essential non-zero $\alpha_{\rho\sigma,\mu\nu}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$:

$$\left. \begin{aligned} \alpha_{11,11}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= a, \\ \alpha_{22,22}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{33,33}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = b, \\ \alpha_{22,33}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{33,22}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = c, \\ \alpha_{11,22}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{11,33}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = \alpha_{22,11}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = \alpha_{33,11}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = d, \\ \alpha_{22,23}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{22,32}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = e, \\ \alpha_{31,31}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{21,21}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = f, \\ \alpha_{12,12}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) &= \alpha_{13,13}\left(\begin{smallmatrix} 1 \\ 12 \end{smallmatrix}\right) = g. \end{aligned} \right\} \quad (6.3)$$

Those $\alpha_{\rho\sigma,\mu\nu}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$ which are obtained by interchanging the 2 and 3 axes have been written out in full; all others may be obtained by the use of equations (6.1), (6.2) and (2.5) and by cyclic interchange of the axes.

The $\alpha\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)$ from (4.5) have now to be calculated. For this purpose the whole set of eigenvectors $e_\rho\left(k\left|\begin{smallmatrix} Q \\ j \end{smallmatrix}\right.\right)$ should be known. These are not given by Kellermann, and even if they were, the calculations for all points in the Q -space would be much too complicated. Now as we have remarked before the integral (5.15) depends essentially on those points of the Q -space where the density of frequency is a maximum. For a linear lattice it can be said that this is the case at the centre point of the cell in reciprocal space. In the three-dimensional case the density $z_{jj'}(\omega)$ of the frequency $\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)$ is defined by the integral

$$z_{jj'}(\omega) d\omega = \iiint_{\omega < \left|\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)\right| < \omega + d\omega} dQ_1 dQ_2 dQ_3. \quad (6.4)$$

The functions $\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)$ are stationary at the centre of the reciprocal cell, $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$;

hence no great error will arise if this value of Q be taken in calculating the $\alpha_{\rho\sigma} \left(\frac{Q}{jj'} \right)$ and the other factors in (5.14). This means that using this approximation (5.15) is replaced by

$$i_{\rho\sigma, \mu\nu}(\omega) = \sum_{jj'} i_{\rho\sigma, \mu\nu}(jj') z_{jj'}(\omega), \quad (6.5)$$

where (jj') now refers to $\left(\frac{Q}{jj'} \right)$ for the point $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

Then the total intensity is

$$I(\omega) = \sum_{\rho\sigma} \sum_{\mu\nu} i_{\rho\sigma, \mu\nu}(\omega) \overline{A_\rho A_\mu^*} \overline{q_\sigma q_\nu}. \quad (6.6)$$

Now consider the case where an unpolarized incident beam along a crystal axis is observed normal to its direction without analyser. Using the formula (1.5) we have put $\theta_1 = 0$, $\theta_2 = \theta_3 = \frac{1}{2}\pi$ and $\phi_2 = 0$, $\phi_1 = \phi_3 = \frac{1}{2}\pi$. Then

$$\left. \begin{aligned} \overline{A_\rho A_\sigma^*} &= 0, \quad \overline{q_\rho q_\sigma} = 0; \quad \rho \neq \sigma, \\ \overline{A_1 A_1^*} &= 0; \quad \overline{A_2 A_2^*} = \overline{A_3 A_3^*} = I_0, \\ \overline{q_1 q_1} &= \overline{q_3 q_3} = \frac{1}{2}; \quad \overline{q_2 q_2} = 0, \end{aligned} \right\} \quad (6.7)$$

giving from (6.6)

$$I(\omega) = \frac{1}{2} I_0 \{ i_{23, 23}(\omega) + i_{31, 31}(\omega) + i_{12, 12}(\omega) + i_{33, 33}(\omega) \}. \quad (6.8)$$

The first three terms are cyclic in the three indices and therefore represent an isotropic behaviour. The last term depends on the axis normal to the plane of the incident and observed light and represents anisotropy around the incident beam. This effect is neglected and $i_{33, 33}(\omega)$ is replaced by $\frac{1}{2} \sum_{\rho} i_{\rho\rho, \rho\rho}(\omega)$, which is a reasonable approximation for cubic crystals. Hence using the fact that in (6.8) the pair $\rho\sigma$ is always the same as the pair $\mu\nu$, then

$$A(jj') = \frac{1}{2} \sum_{\mu\nu} | \alpha_{\mu\nu}(jj') |^2 + \frac{1}{2} \sum_{\mu} | \alpha_{\mu\mu}(jj') |^2 \quad (6.9)$$

must be evaluated.

A lengthy calculation gives the sums (4.5) at the point $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$:

$$\left. \begin{aligned} \alpha_{11}(jj') &= -2F \left\{ \frac{1}{m_1} e_1(1|j) e_1^*(1|j') + \frac{1}{m_2} e_1(2|j) e_1^*(2|j') \right\} \\ &\quad - 2G \left\{ \frac{1}{m_1} [e_2(1|j) e_2^*(1|j') + e_3(1|j) e_3^*(1|j')] \right. \\ &\quad \left. + \frac{1}{m_2} [e_2(2|j) e_2^*(2|j') + e_3(2|j) e_3^*(2|j')] \right\}, \\ \alpha_{23}(jj') &= -2H \left\{ \frac{1}{m_1} [e_2(1|j) e_3^*(1|j') + e_3(1|j) e_2^*(1|j')] \right. \\ &\quad \left. + \frac{1}{m_2} [e_2(2|j) e_3^*(2|j') + e_3(2|j) e_2^*(2|j')] \right\}, \end{aligned} \right\} \quad (6.10)$$

where the constants of (6.3) can now be replaced by three constants

$$F = a + 2b, \quad G = c + 2d, \quad H = e + f + g. \quad (6.11)$$

These three constants determine the intensities of the second-order Raman effect. They are the only arbitrary factors in the expression (6.8) for $I(\omega)$, and are the only quantities which have been adapted to fit the measurements.

In order to evaluate $e_\rho(k|j)$ equations of motion must be established at $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. It can be shown by simple symmetry considerations, as in Kellermann (1940), that the six equations (3.6) split into two sets of three. The first equation of either set is of the form

$$(A - m_k \omega^2) e_1(k|j) + B[e_2(k|j) + e_3(k|j)] = 0, \quad (6.12)$$

with two others obtained by cyclic interchange of the suffix ρ of $e_\rho(k|j)$. The two sets are distinguished by $k = 1$ or 2 . The values of j shall be chosen to correspond with Kellermann's notation. With the help of the orthogonality relations (3.11) the solutions of these equations are

$$\left. \begin{aligned} j = 1: \quad m_1 \omega^2(1) &= A + 2B; \quad e_1(1|j) = e_2(1|j) = e_3(1|j) = \frac{1}{\sqrt{3}}; \\ &e_1(2|j) = e_2(2|j) = e_3(2|j) = 0; \\ j = 2: \quad m_2 \omega^2(2) &= A + 2B; \quad e_1(1|j) = e_2(1|j) = e_3(1|j) = 0; \\ &e_1(2|j) = e_2(2|j) = e_3(2|j) = \frac{1}{\sqrt{3}}. \\ j = 3 \Big| \quad m_1 \omega^2(3) &= m_1 \omega^2(4) = A - B; \quad e_1(1|j) + e_2(1|j) + e_3(1|j) = 0; \\ j = 4 \Big| \quad e_1^2(1|j) + e_2^2(1|j) + e_3^2(1|j) &= 1; \quad e_1(2|j) = e_2(2|j) = e_3(2|j) = 0. \\ j = 5 \Big| \quad m_2 \omega^2(5) &= m_2 \omega^2(6) = A - B; \quad e_1(1|j) = e_2(1|j) = e_3(1|j) = 0; \\ j = 6 \Big| \quad e_1(2|j) + e_2(2|j) + e_3(2|j) &= 0; \quad e_1^2(2|j) + e_2^2(2|j) + e_3^2(2|j) = 1. \end{aligned} \right\} \quad (6.13)$$

Using equations (6.10) in (6.9) and substituting the above values of $e_\rho(k|j)$ after some calculations the following values of $A(jj')$ are found:

$$\left. \begin{aligned} j = j' = 1: \quad A(11) &= \frac{4}{3m_1^2} [\frac{1}{3}(F + 2G)^2 + 4H^2], \\ j = j' = 3 \text{ or } 4: \quad A(33) = A(44) &= \frac{4}{3m_1^2} [\frac{1}{2}(F^2 + 2FG + 3G^2) + \frac{2}{3}H^2], \\ j = 1, j' = 3 \text{ or } 4: \quad A(13) = A(14) &= \frac{4}{3m_1^2} [\frac{1}{3}(F - G)^2 + H^2], \\ j = 3, j' = 4: \quad A(34) &= \frac{1}{2}A(13). \end{aligned} \right\} \quad (6.14)$$

The $A(jj')$ for the $j = 2, 5$ or 6 group are found by replacing m_1 by m_2 . If j belongs to the $(1, 3, 4)$ group and j' to the $(2, 5, 6)$, it can be seen from (6.13) that all products $e_\rho(k|j) e_\sigma^*(k|j')$ are zero; so the corresponding $A(jj')$ is zero.

Combining these results it is found that the approximation to the intensity distribution is

$$I(\omega) = \frac{1}{2} I_0 \sum_{jj'} z_{jj'}(\omega) \mathcal{C}_j \mathcal{C}_{j'} A(jj') \begin{cases} 1 & \text{St.} \\ e^{-\beta_j - \beta_{j'}} & \text{A. St.} \end{cases} \quad (6.15)$$

where \mathcal{C}_j and β_j are taken at the point $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

7. NUMERICAL CALCULATIONS

First calculate the density $z_{jj'}(\omega)$ for each pair of branches. Kellermann gives a table of frequencies of each branch from which $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \pm \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right)$ can be calculated. We found the difficulty that the frequencies collected in the columns $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right)$ are apparently not all of the same geometrical branch, as may be seen by comparing Kellermann's tables with his figures. This is particularly the case for the three branches $\omega\left(\begin{smallmatrix} Q \\ 2 \end{smallmatrix}\right)$, $\omega\left(\begin{smallmatrix} Q \\ 3 \end{smallmatrix}\right)$ and $\omega\left(\begin{smallmatrix} Q \\ 4 \end{smallmatrix}\right)$. But it does not matter greatly as these branches do not differ much. We have taken means between the 3 and 4 branches and between the 5 and 6 branches corresponding to the results for the central point.

The density for $2\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right)$ is taken from Kellermann's paper 'On the specific heat of sodium chloride crystal', where he gives drawings for the density of $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right)$ (Kellermann 1941). Following his methods for the other cases the ω -scale is divided into equal sections $\Delta\omega = 0.6 \times 10^{13} \text{ sec.}^{-1}$ and the number of frequencies $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \pm \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right)$ counted in each interval. By changing the initial point by $0.2 \times 10^{13} \text{ sec.}^{-1}$ three step curves are obtained for each $\omega\left(\begin{smallmatrix} Q \\ jj' \end{smallmatrix}\right)$, which are finally replaced by one smooth curve. The results for the pairs of branches are represented in figure 1.†

One sees already that apart from one group of maxima near zero frequency which will be covered by the width of the exciting mercury line there appear five main groups of maxima. Their mean positions are given in the first row of table 3 and are compared with the peaks observed by Krishnan. The table shows that our theory gives without any arbitrary assumptions the main features of the observed spectrum. A number of small side maxima are to be expected, just as they are found in the photograph, but it would be deceptive to try to include all these in the resultant intensity curve and to compare them with observations, since there is no evidence

† We are indebted to Miss Joan Smith for the calculations of the density distributions of $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \pm \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right)$.

that the smaller peaks in the photograph are not chance fluctuations. Therefore we have used the methods described before, namely, the simplified formula (6.5) containing only twelve combinations of branches.

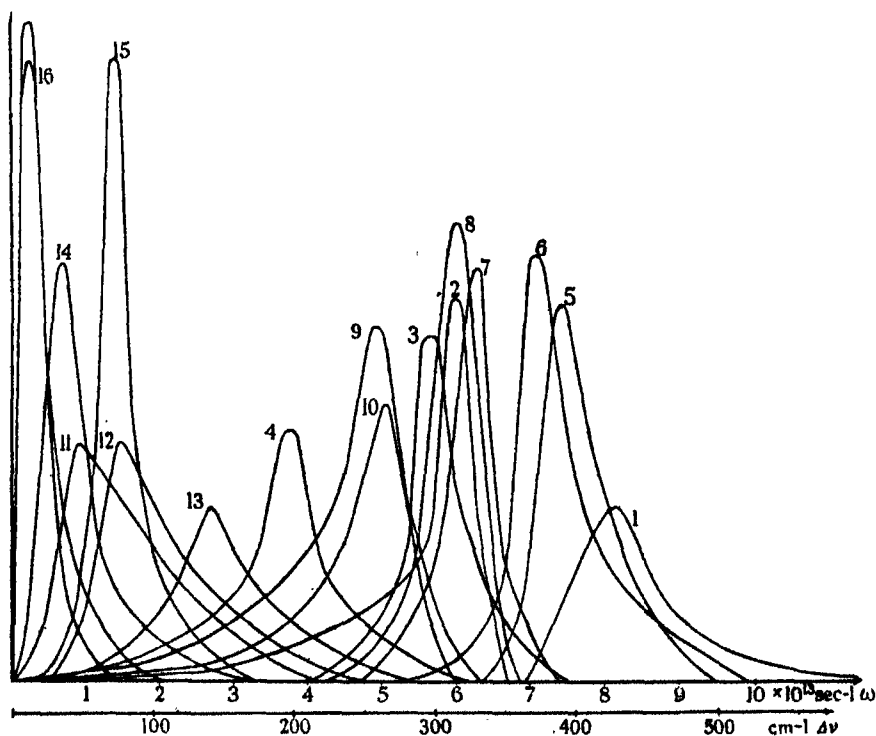


FIGURE 1. Graphs of $z_{j'}(\omega)$, density of frequency $\omega\left(\begin{smallmatrix} Q \\ j \end{smallmatrix}\right) \pm \omega\left(\begin{smallmatrix} Q \\ j' \end{smallmatrix}\right)$.

1. $2\omega(1)$.	7. $\omega(1) + \omega(5), \omega(1) + \omega(6)$.	13. $\omega(1) - \omega(5), \omega(1) - \omega(6)$.
2. $2\omega(2)$.	8. $\omega(2) + \omega(3), \omega(2) + \omega(4)$.	14. $\omega(2) - \omega(3), \omega(2) - \omega(4)$.
3. $2\omega(3), 2\omega(4), \omega(3) + \omega(4)$.	9. $\omega(2) + \omega(5), \omega(2) + \omega(6)$.	15. $\omega(2) - \omega(5), \omega(2) - \omega(6)$.
4. $2\omega(5), 2\omega(6), \omega(5) + \omega(6)$.	10. $\omega(3) + \omega(5), \omega(3) + \omega(6)$.	16. $\omega(3) - \omega(5)$.
5. $\omega(1) + \omega(2)$.	11. $\omega(4) + \omega(5), \omega(4) + \omega(6)$.	17. $\omega(3) - \omega(5), \omega(3) - \omega(6)$,
6. $\omega(1) + \omega(3), \omega(1) + \omega(4)$.	12. $\omega(1) - \omega(2)$.	$\omega(4) - \omega(5), \omega(4) - \omega(6)$.
		18. $\omega(5) - \omega(6)$.

TABLE 3

	$\Delta\nu$ in cm^{-1}					
density maxima from figure 1	140	210	—	265	320	375
observed peaks (Krishnan)	185	184	235	278	314	350

Table 4 contains the numerical values of \mathcal{G}_j and $e^{-\beta_j}$ for $Q = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $T = 273^\circ \text{K}$. Using products and squares of \mathcal{G}_j and the corresponding $e^{-\beta_j}$ according to formula (6.15) we have redrawn the branches after multiplication by the proper factors,

TABLE 4

	$\omega(1)$	$\omega(2)$	$\omega(3) = \omega(4)$	$\omega(5) = \omega(6)$
$\frac{2\mathcal{C}_j}{h} \times 10^{13}$	0.309	0.429	0.855	1.245
$e^{-\beta_j}$	0.284	0.361	0.512	0.584

obtaining a figure which differs only a little from figure 1. Finally, we have tried to fit the maxima to the observed intensities by adapting the $A(jj')$ of (6.14) which depend on the three constants F , G and H . We found that the simplest relations between F , G and H leading to fairly good relative intensities were

$$F + G = 0, \quad H = 0, \quad (7.1)$$

$$\text{giving from (6.14)} \quad A(11):A(34):A(33):A(13) = 1:2:3:4. \quad (7.2)$$

Considering the number of omissions and simplifications which we have already made it does not seem worth while to improve the results by changing the constants $A(jj')$. It should be possible to determine the coupling constants from first principles by applying quantum mechanics to the interaction between the electronic clouds of neighbouring atoms. We should like to draw attention to the fact that one of the constants F or G appears to be negative (probably G).

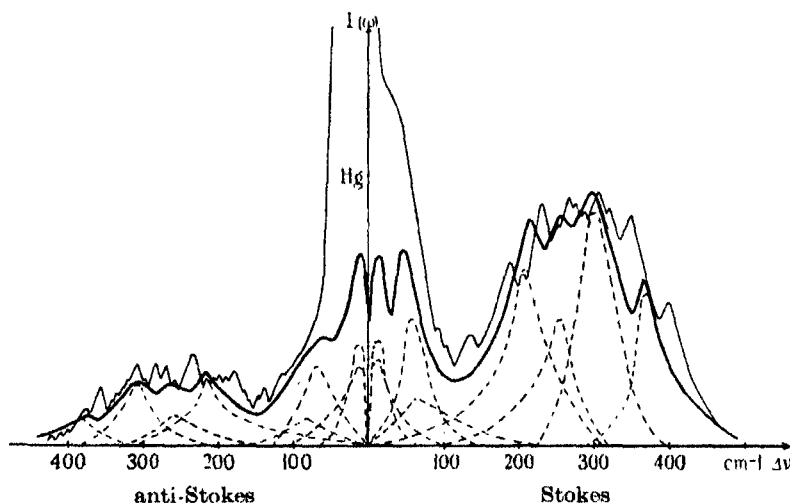


FIGURE 2. Fine line is sketch of Krishnan's microphotometric record of Raman spectrum of rock-salt (Krishnan 1945*a*). Broken lines represent theoretical contributions to intensity of pairs of branches of frequency. Thick line represents theoretical intensity obtained from superposition of these.

Figure 2 shows the final result. It contains the contributions of the single branches, their superposition and the experimental curve taken from Krishnan's microphotograph (Krishnan 1945*a*). The general features of the observations are well

represented. If one takes into account that many of the curves are actually double or triple and that some have been entirely omitted, one would expect that a more perfect approximation would give an even better agreement.†

Our results show conclusively that the second-order Raman effect cannot be used to disprove lattice dynamics. The observations not only do not confirm any of the predictions of Raman's theory but they represent direct evidence for the existence and quantitative properties of the continuous spectrum as predicted by correct mechanical treatment of lattice vibrations.

REFERENCES

- Blackman, M. 1933 *Z. Phys.* **86**, 421.
Born, M. 1946 *Nature*, **157**, 810.
Born, M. & Blackman, M. 1933 *Z. Phys.* **82**, 551.
Born, M. & Bradburn, M. 1945 *Nature*, **156**, 567.
Born, M. & Oppenheimer, R. 1927 *Ann. Phys., Lpz.*, **84**, 551.
Fermi, E. & Rasetti, F. 1931 *Z. Phys.* **71**, 689.
Kellermann, E. W. 1940 *Phil. Trans. A*, **238**, 513.
Kellermann, E. W. 1941 *Proc. Roy. Soc. A*, **178**, 17.
Krishnan, R. S. 1943 *Proc. Ind. Acad. Sci. A*, **18**, 298.
Krishnan, R. S. 1945*a* *Nature*, **156**, 267.
Krishnan, R. S. 1945*b* *Nature*, **155**, 171.
Krishnan, R. S. 1946 *Nature*, **157**, 623.
Placzek, G. 1934 Marx, *Handbuch der Radiologie*, **6**, part 2.
Rasetti, F. 1931 *Nature*, **127**, 626.

† The figure published in our preliminary publication in *Nature* (1945; see also Krishnan 1946 and Born 1946) differs a little from figure 2 of the text because of a numerical mistake; we have given one of the branches, $\omega(5)$, double its weight.

Thermal scattering of X-rays by crystals

I. Dynamical foundation

BY G. H. BEGBIE AND M. BORN, F.R.S., *University of Edinburgh*

(Received 9 October 1945)

The dynamical properties of a crystal for small vibrations can be described by the set of coefficients of the potential energy forming the dynamical matrix. The elastic constants and many other observable quantities can be calculated in terms of the elements of the dynamical matrix (but in general not vice versa). The formulae given in previous publications are based on too narrow assumptions (central forces). In the present paper (Part I) this theory is developed in the most general form, without any assumption about the nature of the atomic forces. It contains, further, the method for reducing the dynamical matrix with the help of the space group of the lattice.

INTRODUCTION

The purpose of this paper is the calculation of the thermal scattering of X-rays for a few special lattices. The general theory of this effect has been given by Waller (1925) and by several other physicists. A condensed presentation of this theory appeared in *Reports on Progress in Physics* (Born 1942-43), which contains a bibliography of the subject and an account of its development. The main point of this *Report* is that the scattering power of a crystal is not expressed in terms of the frequencies but directly in terms of the 'dynamical matrix' of the lattice, which is defined as the system of the coefficients of the potential energy for small displacements of the particles.

In the present paper we intend to calculate the scattering without assuming particular laws of force between the particles. The potential energy between the nuclei, defined as the lowest eigenvalue of the electronic energy for fixed nuclei in positions near to the regular spacing of the lattice, is assumed to be any arbitrary function of the displacements of the nuclei. All physical properties can be expressed in terms of this function and its derivatives. The vibrational properties depend in particular on the second derivatives which form the dynamical matrix. For practical purposes we shall make the single assumption that the elements of this matrix are negligibly small except for next neighbours. Now in the existing literature, in particular the book *Atomtheorie des festen Zustandes* (Born 1923), the dynamics of the lattice has not been developed in such a general way, but only under the assumption of central forces. In an older book, *Dynamik der Kristallgitter* (Born 1915), an attempt was made to work with the general potential energy restricted only by the necessary invariance against translation and rotation; but the results obtained are not quite correct. Therefore it was necessary to reconsider the problem of deriving the macroscopic properties of a crystal from the potential energy function. The

general theory will be published later (in book form). Here we give only the connexion of the elastic constants with the dynamical matrix and the equation for elastic waves.

The present first part contains, further, the general symmetry considerations by which the dynamical matrix is simplified for a given special lattice. Application to some lattices will be treated in the second part.

1. LATTICE DYNAMICS

A short summary of the principal definitions, as used in the *Report* quoted above, is given first.

Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ be the three elementary vectors of the lattice cell, and $\mathbf{r}(k)$ ($k = 1, 2, \dots, n$) the position vectors of the n particles (nuclei) in the cell. Then an arbitrary lattice vector shall be written

$$\mathbf{r}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = \mathbf{r}(l) + \mathbf{r}(k), \quad (1.1)$$

$$\text{where} \quad \mathbf{r}(l) = l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3 \quad (l_1, l_2, l_3 \text{ are integers}). \quad (1.2)$$

Let the rectangular components of $\mathbf{r}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$ be $x_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$ ($\alpha = 1, 2, 3$). We use these symbols also if the particles are a little displaced from their lattice positions; the latter are, if necessary, characterized by an index 0. The potential energy for a disturbed but still lattice-like state will be denoted by Φ which is a function of all the $x_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$. According to a well-known theorem of quantum mechanics (Born & Oppenheimer 1927) it is the lowest eigenvalue of the total electronic energy, if the nuclei are fixed in the positions $x_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$.

The second derivatives of Φ in equilibrium which depend only on $l-l'$ (not on l and l' separately) will be denoted by

$$\Phi_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) = \left[\frac{\partial^2 \Phi}{\partial x_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) \partial x_\beta\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix}\right)} \right]_0. \quad (1.3)$$

$$\text{They satisfy the identity} \quad \Phi_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) = \Phi_{\beta\alpha}\left(\begin{smallmatrix} -l \\ k'k \end{smallmatrix}\right). \quad (1.4)$$

Then the equations of motion for small vibrations, of amplitude

$$u_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = x_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) - x_\alpha^0\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right),$$

$$\text{are} \quad m_k \ddot{u}_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) + \sum_{\beta l' k'} \Phi_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) u_\beta\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix}\right) = 0, \quad (1.5)$$

where m_k is the mass of the particle of the type k .

The quantities $\Phi_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$ represent the quasi-elastic forces between any point $\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$ and a point of the base $\left(\begin{smallmatrix} 0 \\ k' \end{smallmatrix}\right)$. In particular, the force of a point $\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)$ on itself, $\Phi\left(\begin{smallmatrix} 0 \\ kk \end{smallmatrix}\right)$, can be determined from all the other forces, for if the whole lattice is subjected to a translation, $u_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = \text{const.}$, $\ddot{u}_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = 0$, then from (1.5) one obtains

$$\sum_{l'k'} \Phi_{\alpha\beta}\left(\begin{smallmatrix} l' \\ kk' \end{smallmatrix}\right) = 0, \quad \text{or} \quad \Phi_{\alpha\beta}\left(\begin{smallmatrix} 0 \\ kk \end{smallmatrix}\right) = - \sum_{l'k'}' \Phi_{\alpha\beta}\left(\begin{smallmatrix} l' \\ kk' \end{smallmatrix}\right), \quad (1.6)$$

where a dash on the summation sign denotes that the terms in which $l' = 0$, $k = k'$ is to be omitted from the summation. Further identities of a similar kind follow from the fact that the potential energy is invariant with respect to rotations. By applying an infinitesimal rotation one finds, after some calculation,

$$\left. \begin{aligned} \sum_{lk} \Phi_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) x_{\lambda}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) &= \sum_{lk} \Phi_{\lambda\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) x_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right), \\ \sum_{l'k'} \Phi_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) x_{\mu}\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix}\right) &= \sum_{l'k'} \Phi_{\alpha\mu}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) x_{\beta}\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix}\right). \end{aligned} \right\} \quad (1.7)$$

One can get rid of the mass factor in (1.5) by introducing the reduced amplitudes

$$v_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = \sqrt{m_k} u_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right), \quad (1.8)$$

and the coefficients
$$D_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) = \frac{1}{\sqrt{(m_k m_{k'})}} \Phi_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right), \quad (1.9)$$

which satisfy
$$D_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) = D_{\beta\alpha}\left(\begin{smallmatrix} -l \\ k'k \end{smallmatrix}\right). \quad (1.10)$$

Then the equations of motion are

$$\ddot{v}_{\alpha}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = - \sum_{\beta l'k'} D_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right) v_{\beta}\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix}\right). \quad (1.11)$$

$D_{\alpha\beta}\left(\begin{smallmatrix} l-l' \\ kk' \end{smallmatrix}\right)$, considered as a function of the two index-triples (α, l, k) and (β, l', k') , is an element of the dynamical matrix, which can be decomposed into submatrices $D(l-l')$, where

$$D(l) = \begin{pmatrix} D\left(\begin{smallmatrix} l \\ 11 \end{smallmatrix}\right) & D\left(\begin{smallmatrix} l \\ 12 \end{smallmatrix}\right) & \dots & D\left(\begin{smallmatrix} l \\ 1n \end{smallmatrix}\right) \\ D\left(\begin{smallmatrix} l \\ 21 \end{smallmatrix}\right) & D\left(\begin{smallmatrix} l \\ 22 \end{smallmatrix}\right) & \dots & D\left(\begin{smallmatrix} l \\ 2n \end{smallmatrix}\right) \\ \dots & \dots & \dots & \dots \\ D\left(\begin{smallmatrix} l \\ n1 \end{smallmatrix}\right) & D\left(\begin{smallmatrix} l \\ n2 \end{smallmatrix}\right) & \dots & D\left(\begin{smallmatrix} l \\ nn \end{smallmatrix}\right) \end{pmatrix}. \quad (1.12)$$

Here each element is again a (3×3) matrix

$$D\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) = \begin{pmatrix} D_{11}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) & D_{12}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) & D_{13}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) \\ D_{21}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) & \dots\dots\dots & \dots\dots\dots \\ D_{31}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) & \dots\dots\dots & \dots\dots\dots \end{pmatrix}. \quad (1.13)$$

Let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be the three elementary vectors of the reciprocal lattice so that

$$\mathbf{a}_\alpha \cdot \mathbf{b}_\beta = \delta_{\alpha\beta} \quad \left(\delta_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \neq \beta \end{cases} \right), \quad (1.14)$$

and let

$$\mathbf{q} = \mathbf{b}_1 Q_1 + \mathbf{b}_2 Q_2 + \mathbf{b}_3 Q_3 \quad (1.15)$$

be the position vector of a point in the reciprocal space. If a vector \mathbf{r} be written in the real space in the form

$$\mathbf{r} = \mathbf{a}_1 X_1 + \mathbf{a}_2 X_2 + \mathbf{a}_3 X_3, \quad (1.16)$$

then clearly

$$\mathbf{r} \cdot \mathbf{q} = X_1 Q_1 + X_2 Q_2 + X_3 Q_3 = (XQ). \quad (1.17)$$

Hence, if \mathbf{r} is a lattice-point $\mathbf{r}(l)$, i.e. $X_\alpha = l_\alpha$ (an integer), and \mathbf{q} is a point of the reciprocal lattice, i.e. $Q_\alpha = h_\alpha$ (an integer), then the scalar product $\mathbf{r} \cdot \mathbf{q}$ is an integer. Hence $e^{2\pi i \mathbf{q} \cdot \mathbf{r}(l)} = e^{2\pi i (Ql)}$ is periodic in the reciprocal lattice.

One can now construct the Fourier transform of the dynamical matrix, or its representation in reciprocal space. As shown in the *Report* (Born 1942-43) this matrix is diagonal with respect to the \mathbf{q} or (Q) ; one has

$$D_{\alpha\beta}(\mathbf{q}\mathbf{q}' | kk') = \delta(\mathbf{q}\mathbf{q}') D_{\alpha\beta}(\mathbf{q} | kk'), \quad (1.18)$$

where
$$D_{\alpha\beta}(\mathbf{q} | kk') = \sum_l D_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) e^{-2\pi i \mathbf{q} \cdot \mathbf{r}(l)} = \sum_l D_{\alpha\beta}\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right) e^{-2\pi i (lQ)}. \quad (1.19)$$

The quantities (1.19) will next be considered as elements of the matrix $D(\mathbf{q})$, with the pairs (αk) and $(\beta k')$ as row and column indices.

The equations of motion (1.11) can now be solved by plane waves which can be written in the two forms:

$$v_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = V_\alpha(k) e^{-i\omega t} e^{2\pi i \mathbf{q} \cdot \mathbf{r}(l)}, \quad (1.20)$$

and

$$v_\alpha\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right) = W_\alpha(k) e^{-i\omega t} e^{2\pi i \mathbf{q} \cdot \mathbf{r}\left(\begin{smallmatrix} l \\ k \end{smallmatrix}\right)}. \quad (1.21)$$

Then clearly from (1.1)

$$W_\alpha(k) = V_\alpha(k) e^{-2\pi i \mathbf{q} \cdot \mathbf{r}(k)}. \quad (1.22)$$

For each \mathbf{q} separately, the amplitudes $V_\alpha(k)$ and $W_\alpha(k)$ satisfy respectively the equations

$$\omega^2 V_\alpha(k) = \sum_{\beta k'} D_{\alpha\beta}(\mathbf{q} | kk') V_\beta(k'), \quad (1.23)$$

$$\omega^2 W_\alpha(k) = \sum_{\beta k'} C_{\alpha\beta}(\mathbf{q} | kk') W_\beta(k'), \quad (1.24)$$

where $C_{\alpha\beta}(\mathbf{q} | kk') = D_{\alpha\beta}(\mathbf{q} | kk') e^{-2\pi i \mathbf{q} \cdot \mathbf{r}(kk')} = \sum_l D_{\alpha\beta} \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) e^{-2\pi i \mathbf{q} \cdot \mathbf{r} \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right)}, \quad (1.25)$

with

$$\mathbf{r}(kk') = \mathbf{r}(k) - \mathbf{r}(k').$$

We introduce the matrix $C(\mathbf{q})$, with row and column suffixes formed from the index pairs (αk) and $(\beta k')$ respectively. If $\epsilon(\mathbf{q})$ is the diagonal matrix with the element

$$\epsilon_{\alpha\beta}(\mathbf{q} | kk') = \delta_{\alpha\beta} \delta_{kk'} e^{-2\pi i \mathbf{q} \cdot \mathbf{r}(k)}, \quad (1.26)$$

one can write (1.25) $C(\mathbf{q}) = \epsilon(\mathbf{q}) D(\mathbf{q}) \epsilon^*(\mathbf{q}). \quad (1.27)$

The solution with $V_\alpha(k)$ is convenient for introducing normal co-ordinates, but that with $W_\alpha(k)$ is more suitable for discussing the properties of the waves, particularly in the case of long wave-length.

2. LONG WAVES

The equations of motion are written in the compact form

$$\Omega W = C(\mathbf{q}) W, \quad (2.1)$$

where $\Omega = \omega^2$ and W is a one-column matrix:

$$W = \{W_1(1), W_2(1), W_3(1), W_1(2), \dots, W_2(n), W_3(n)\}. \quad (2.2)$$

The analysis that follows has already been given by Born (1923) in another notation.

Now consider the matrix $C(\mathbf{q})$ as expanded in powers of the wave-number, thus

$$C(\mathbf{q}) = C^{(0)} + C^{(1)}(\mathbf{q}) + C^{(2)}(\mathbf{q}) + \dots, \quad (2.3)$$

where the elements of $C^{(0)}$ are constant, those of $C^{(1)}(\mathbf{q})$ linear in \mathbf{q} , etc., namely,

$$C_{\alpha\beta}^{(0)}(kk') = \frac{1}{\sqrt{(m_k m_{k'})}} \sum_l \Phi_{\alpha\beta} \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right), \quad (2.4)$$

$$C_{\alpha\beta}^{(1)}(\mathbf{q} | kk') = \frac{-2\pi i}{\sqrt{(m_k m_{k'})}} \sum_l \sum_\gamma \Phi_{\alpha\beta} \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) x_\gamma \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) \mathbf{q}_\gamma, \quad (2.5)$$

$$C_{\alpha\beta}^{(2)}(\mathbf{q} | kk') = \frac{-(2\pi)^2}{\sqrt{(m_k m_{k'})}} \sum_l \sum_{\lambda\mu} \Phi_{\alpha\beta} \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) x_\lambda \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) x_\mu \left(\begin{smallmatrix} l \\ kk' \end{smallmatrix} \right) \mathbf{q}_\lambda \mathbf{q}_\mu. \quad (2.6)$$

If one now expands Ω and W in the same way,

$$\Omega = \Omega^{(0)} + \Omega^{(1)} + \Omega^{(2)} + \dots, \quad (2.7)$$

$$W = W^{(0)} + W^{(1)} + W^{(2)} + \dots, \quad (2.8)$$

then follows, by comparison of equal powers in (2.1), a set of equations

$$\stackrel{(0)(0)}{\Omega} W = \stackrel{(0)(0)}{C} W, \quad (2.9)$$

$$\stackrel{(0)(1)}{\Omega} W + \stackrel{(1)(0)}{\Omega} W = \stackrel{(0)(1)}{C} W + \stackrel{(1)(0)}{C} W, \quad (2.10)$$

$$\stackrel{(0)(2)}{\Omega} W + \stackrel{(1)(1)}{\Omega} W + \stackrel{(2)(0)}{\Omega} W = \stackrel{(0)(2)}{C} W + \stackrel{(1)(1)}{C} W + \stackrel{(2)(0)}{C} W, \quad (2.11)$$

.....

Consider first the zero approximation (2.9). These equations must be satisfied for any translations of the crystal as a whole. There are three independent translations, namely,

$$\left. \begin{aligned} u_1 \binom{l}{k} &= \frac{1}{\sqrt{m_k}} v_1 \binom{l}{k} = U_1 \quad \text{or} \quad 0 \quad \text{or} \quad 0, \\ u_2 \binom{l}{k} &= \frac{1}{\sqrt{m_k}} v_2 \binom{l}{k} = 0 \quad U_2 \quad 0, \\ u_3 \binom{l}{k} &= \frac{1}{\sqrt{m_k}} v_3 \binom{l}{k} = 0 \quad 0 \quad U_3. \end{aligned} \right\} \quad (2.12)$$

The corresponding $W_\alpha(k)$ are obtained from (1.21) for $\omega = 0$ and $\mathbf{q} = 0$; hence the equations (2.9) have three sets of trivial solutions

$$\left. \begin{aligned} \stackrel{(0)}{W}_1 \binom{l}{k} &= \sqrt{m_k} U_1 \quad \text{or} \quad 0 \quad \text{or} \quad 0, \\ \stackrel{(0)}{W}_2 \binom{l}{k} &= 0 \quad \sqrt{m_k} U_2 \quad 0, \\ \stackrel{(0)}{W}_3 \binom{l}{k} &= 0 \quad 0 \quad \sqrt{m_k} U_3, \end{aligned} \right\} \quad (2.13)$$

the right-hand terms being independent of l .

By using the matrix of $3n$ lines and 3 columns

$$M = \begin{pmatrix} \sqrt{m_1} & 0 & 0 \\ 0 & \sqrt{m_1} & 0 \\ 0 & 0 & \sqrt{m_1} \\ \sqrt{m_2} & 0 & 0 \\ \dots\dots\dots \\ 0 & 0 & \sqrt{m_n} \end{pmatrix}; \quad (2.14)$$

this solution (2.13) can be written in the condensed form

$$\stackrel{(0)}{\Omega} = 0, \quad \stackrel{(0)}{W} = MU, \quad (2.15)$$

where

$$U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}. \quad (2.16)$$

Consider next the first-order equation (2.10), and write it in the form

$$^{(0)(1)}CW = ^{(1)}\Omega MU - ^{(1)}C(\mathbf{q}) MU. \quad (2.17)$$

Since the homogeneous equations $^{(0)(1)}CW = 0$ have a solution of the form $^{(1)}W = MU'$, where U' is an arbitrary one-column matrix like U in (2.16), the equations (2.17) are only solvable if

$$^{(1)}\tilde{M}MU - \tilde{M}^{(1)}C(\mathbf{q}) MU = 0, \quad (2.18)$$

where \tilde{M} is the transpose of the matrix M .

Now from (2.5), together with (1.6), it follows, by writing out the elements, that

$$\tilde{M}^{(1)}C(\mathbf{q}) M = 0. \quad (2.19)$$

Hence $^{(1)}\Omega = 0, \quad (2.20)$

and (2.17) reduces to $^{(0)(1)}CW = -^{(1)}C(\mathbf{q}) MU. \quad (2.21)$

The solution of this equation has the form

$$^{(1)}W = -^{(0)(1)}\Gamma C(\mathbf{q}) MU + MU', \quad (2.22)$$

where MU' is the solution of the homogeneous equations.

Now the second-order equation (2.11) gives

$$^{(0)(2)}CW = ^{(2)}\Omega MU - ^{(1)}C(\mathbf{q}) ^{(1)}W - ^{(2)}C(\mathbf{q}) MU. \quad (2.23)$$

The solvability condition is

$$^{(2)}\tilde{M}MU = \tilde{M}^{(1)}C(\mathbf{q}) ^{(1)}W + \tilde{M}^{(2)}C(\mathbf{q}) MU. \quad (2.24)$$

$^{(1)}W$ can now be eliminated between the two equations (2.22) and (2.24); the term with U' vanishes because of (2.19) and the term $\tilde{M}^{(0)}C(\mathbf{q}) ^{(3)}W$ because of (1.6); hence

$$^{(2)}\tilde{M}MU = \tilde{M}\{^{(2)}C(\mathbf{q}) - ^{(1)}C(\mathbf{q}) ^{(0)(1)}\Gamma C(\mathbf{q})\} MU. \quad (2.25)$$

As the expression in brackets is of the second order in the components of \mathbf{q} , the explicit expression of (2.25) has the form

$$\rho\omega^2 U_\alpha = \sum_{\beta} \sum_{\lambda\mu} \{\alpha\lambda, \beta\mu\} \mathbf{q}_\lambda \mathbf{q}_\mu U_\beta, \quad (2.26)$$

where $\rho = \frac{1}{v_\alpha} \sum_k m_k$ is the density, v_α being the volume of the unit cell. By introducing

the values of $^{(1)}C(\mathbf{q})$ and $^{(2)}C(\mathbf{q})$ from (2.5) and (2.6), one gets the expression for the coefficient in (2.26)

$$\{\alpha\lambda, \beta\mu\} = \frac{4\pi^2}{v_\alpha} \sum_l \sum_{kk'} \left\{ \sum_{l'} \sum_{k_1 k_2} \Phi_{\alpha\gamma} \left(\begin{matrix} l \\ kk_1 \end{matrix} \right) x_\lambda \left(\begin{matrix} l \\ kk_1 \end{matrix} \right) \Gamma_{\gamma\delta}^{(0)}(k_1 k_2) \Phi_{\delta\beta} \left(\begin{matrix} l' \\ k_2 k' \end{matrix} \right) x_\mu \left(\begin{matrix} l' \\ k_2 k' \end{matrix} \right) \right. \\ \left. - \frac{1}{2} \Phi_{\alpha\beta} \left(\begin{matrix} l \\ kk' \end{matrix} \right) x_\lambda \left(\begin{matrix} l \\ kk' \end{matrix} \right) x_\mu \left(\begin{matrix} l \\ kk' \end{matrix} \right) \right\}. \quad (2.27)$$

From this there follow, using (1.7), the symmetry properties

$$\{\alpha\lambda, \beta\mu\} = \{\alpha\lambda, \mu\beta\} = \{\lambda\alpha, \beta\mu\} = \{\beta\mu, \alpha\lambda\}. \quad (2.28)$$

One can now compare (2.26) with the equation for the amplitudes of elastic waves in ordinary elasticity theory. They have exactly the same form, where $\{\alpha\lambda, \beta\mu\}$ are the elastic constants. In virtue of the identities (2.28) one can introduce Voigt's notation, replacing the index combinations

$$\alpha\lambda = 11 \quad 22 \quad 33 \quad 23 \quad 31 \quad 12$$

by

$$\rho = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$$

Then $\{\alpha\lambda, \beta\mu\}$ can be simply written $c_{\rho\sigma}$, with $c_{\rho\sigma} = c_{\sigma\rho}$, and (2.26) assumes the form

$$\rho\omega^2 U_\alpha = \sum_\beta D'_{\alpha\beta}(\mathbf{q}) U_\beta, \quad (2.29)$$

where the elements of $D'(\mathbf{q})$ are given by the matrix equation

$$\begin{pmatrix} D'_{11}(\mathbf{q}) \\ D'_{22}(\mathbf{q}) \\ D'_{33}(\mathbf{q}) \\ D'_{23}(\mathbf{q}) \\ D'_{31}(\mathbf{q}) \\ D'_{12}(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} c_{11} & c_{66} & c_{55} & c_{65} & c_{51} & c_{16} \\ c_{66} & c_{22} & c_{44} & c_{24} & c_{46} & c_{62} \\ c_{55} & c_{44} & c_{33} & c_{43} & c_{35} & c_{54} \\ c_{65} & c_{24} & c_{43} & \frac{1}{2}(c_{23} + c_{44}) & \frac{1}{2}(c_{45} + c_{36}) & \frac{1}{2}(c_{64} + c_{25}) \\ c_{51} & c_{46} & c_{35} & \frac{1}{2}(c_{45} + c_{36}) & \frac{1}{2}(c_{31} + c_{55}) & \frac{1}{2}(c_{56} + c_{14}) \\ c_{16} & c_{62} & c_{54} & \frac{1}{2}(c_{64} + c_{25}) & \frac{1}{2}(c_{56} + c_{14}) & \frac{1}{2}(c_{12} + c_{66}) \end{pmatrix} \begin{pmatrix} q_1^2 \\ q_2^2 \\ q_3^2 \\ 2q_2q_3 \\ 2q_3q_1 \\ 2q_1q_2 \end{pmatrix}. \quad (2.30)$$

The results may be summarized in the following recipe for determining the limiting case of long waves:

Establish (2.21) and (2.24) for a given lattice. Written in component form they are

$$\sum_\beta \sum_{k'}^{(0)} C_{\alpha\beta}(kk') \overset{(1)}{W}_\beta(k') = - \sum_\beta \sum_{k'} \sqrt{m_{k'}} \overset{(1)}{C}_{\alpha\beta}(\mathbf{q} | kk') U_\beta, \quad (2.31)$$

$$\sum_k \overset{(2)}{m}_k \Omega U_\alpha = \sum_{kk'} \sum_\beta \sqrt{m_k} \overset{(1)}{C}_{\alpha\beta}(\mathbf{q} | kk') \overset{(1)}{W}_\beta(k') + \sum_{kk'} \sqrt{(m_k m_{k'})} \overset{(2)}{C}_{\alpha\beta}(\mathbf{q} | kk') U_\beta. \quad (2.32)$$

Then eliminate $\overset{(1)}{W}$. The result has the form of (2.29) and allows the elastic constants to be read off in terms of combinations of the elements of the dynamical matrix.

This method shall be used in Part II for special lattices.

3. SYMMETRY OF THE LATTICE

In order to find the relations that exist between the elements of the dynamical matrix, all those symmetry operations have to be applied to the lattice which will bring it into self-coincidence. The number and nature of these operations are described by the space group of the lattice. They comprise rotations, reflexions, glide planes

and combinations of these. (Translations by multiples of the base vectors have already been taken into account by assuming that the dynamical matrix $D_{kk'}^{(l-l')}$ only depends on the combination $l-l'$, and not on l and l' separately.)

Consider the rectangular system of axes in which a point of space has the co-ordinates x_α ($\alpha = 1, 2, 3$). Let a transformation T act on these axes so that they take up a new position, and in this transformed system let the point have the co-ordinates x_A ($A = 1, 2, 3$). Then relations will hold of the form

$$x_\alpha = \sum_A T_{\alpha A} x_A. \quad (3.1)$$

The coefficients $T_{\alpha A}$ depend on the particular transformation chosen. If the point x_α is taken to be a point of the lattice then (3.1) becomes

$$x_\alpha \begin{pmatrix} l \\ k \end{pmatrix} = \sum_A T_{\alpha A} x_A \begin{pmatrix} l \\ k \end{pmatrix}. \quad (3.2)$$

Let T now be a symmetry operation of the lattice, then it will be possible to find a point of the lattice $\begin{pmatrix} L \\ K \end{pmatrix}$, say, such that

$$x_\alpha \begin{pmatrix} l \\ k \end{pmatrix} = x_B \begin{pmatrix} L \\ K \end{pmatrix} \quad \text{for} \quad \begin{matrix} \alpha \\ B \end{matrix} = 1, 2, 3. \quad (3.3)$$

A combination of (3.2) and (3.3) gives

$$x_B \begin{pmatrix} L \\ K \end{pmatrix} = \sum_A T_{\alpha A} x_A \begin{pmatrix} l \\ k \end{pmatrix}. \quad (3.4)$$

If the symbol T stands for the matrix formed from the quantities $T_{\alpha A}$, $T \equiv (T_{\alpha A})$, and \mathbf{r} is considered as a one-column matrix, then (3.4) can be written in the matrix form

$$\mathbf{r} \begin{pmatrix} L \\ K \end{pmatrix} = T \mathbf{r} \begin{pmatrix} l \\ k \end{pmatrix}. \quad (3.5)$$

The elements $D_{\alpha\beta} \begin{pmatrix} l-l' \\ kk' \end{pmatrix}$ of the dynamical matrix behave, with respect to the indices α and β , as a covariant tensor of rank two. Their transformation law will be

$$D_{\alpha\beta} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} = \sum_{A,B} T_{\alpha A} T_{\beta B} D_{AB} \begin{pmatrix} l-l' \\ kk' \end{pmatrix}, \quad \begin{matrix} A \\ B \end{matrix} = 1, 2, 3. \quad (3.6)$$

Now points $\begin{pmatrix} L \\ K \end{pmatrix}$ and $\begin{pmatrix} L' \\ K' \end{pmatrix}$ can be found such that

$$D_{\alpha\beta} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} = D_{MN} \begin{pmatrix} L-L' \\ KK' \end{pmatrix}, \quad \begin{matrix} \alpha, \beta \\ M, N \end{matrix} = 1, 2, 3. \quad (3.7)$$

Hence
$$D_{MN} \begin{pmatrix} L-L' \\ KK' \end{pmatrix} = \sum_{A,B} T_{\alpha A} T_{\beta B} D_{AB} \begin{pmatrix} l-l' \\ kk' \end{pmatrix}, \quad (3.8)$$

or, in matrix notation,
$$D \begin{pmatrix} L-L' \\ KK' \end{pmatrix} = T D \begin{pmatrix} l-l' \\ kk' \end{pmatrix} \tilde{T}, \quad (3.9)$$

where \tilde{T} is the transpose of the matrix T , and the change from $\begin{pmatrix} l \\ k \end{pmatrix}$ to $\begin{pmatrix} L \\ K \end{pmatrix}$ and from $\begin{pmatrix} l' \\ k' \end{pmatrix}$ to $\begin{pmatrix} L' \\ K' \end{pmatrix}$ is to be obtained from (3.4) or (3.5).

The space group of a given lattice has a base from which all group elements can be obtained by composition. Let T_1, T_2, \dots be the matrices representing this base. If these are introduced in (3.9) one obtains a set of matrix equations from which identities between the elements of the dynamical matrix can be read off.

In Part II these identities will be constructed for special lattices.

We have to thank Dr H. W. Peng for valuable discussions and advice.

REFERENCES

- Born, M. 1915 *Dynamik der Kristallgitter*. Leipzig.
 Born, M. 1923 *Atomtheorie des festen Zustandes*. Leipzig.
 Born, M. 1942-3. *Reports on Progress in Physics*, **9**, 294.
 Born, M. & Oppenheimer, R. 1927 *Ann. Phys., Lpz.*, **84**, 457.
 Waller, I. 1925 Diss. (Uppsala).

Thermal scattering of X-rays by crystals

II. The thermal scattering of the face-centred cubic and the close-packed hexagonal lattices

BY G. H. BEGBIE, *University of Edinburgh*

(Communicated by M. Born, F.R.S.—Received 9 October 1945)

The results of the preceding paper (Part I) are here applied to two special lattices, the face-centred cubic and the close-packed hexagonal lattices. In both cases the assumption is made that only next-neighbour atoms act on one another. In the case of the cubic lattice the number of atomic constants turns out to be equal to that of the elastic constants, so that the dynamical matrix can be expressed in terms of the latter. Numerical calculations are performed taking for the elastic constants those of potassium chloride; the results are compared with those obtained by Iona who used the correct ionic forces. Then the scattering matrix is calculated and a diagram of equi-diffusion lines is drawn which covers a part of the reciprocal space containing nine lattice points.

In the case of the hexagonal lattice the number of atomic constants is greater than that of the elastic constants. The dynamical matrix is given in terms of the former; but equi-diffusion lines are constructed only for the vicinity of the selective reflexions (Jahn case) where the scattering can be expressed in terms of the elastic constants.

INTRODUCTION

With the help of the general formulae and rules established in Part I, the thermal scattering will be calculated for the two special cases mentioned in the title. In the first section a recapitulation of the formulae and notation, as used in *Reports on Progress in Physics* (Born 1942-3) is given. As stated in the introduction to Part I the dynamical matrix is simplified by assuming that only elements belonging to neighbour particles are essentially different from zero. This assumption and proper symmetry considerations (according to Part I, § 4) reduce the number of arbitrary constants considerably. By going over to long waves, in accordance with the rules of § 3 of Part I, the so-called Jahn case is obtained (Jahn 1941-2), which represents the scattering in the neighbourhood of selective reflexions (Laue-Bragg spots) and depends only on the macroscopic elastic constants.

Now it turns out that the number of atomic constants (independent constants in the simplified dynamical matrix) is exactly the same as the number of elastic constants for the face-centred cubic lattice. Hence it is possible to express the dynamical matrix and therefore also the scattering matrix rigorously in terms of the elastic constants, also in regions where the Jahn approximation is not valid. Numerical calculations have been performed with these formulae, covering a whole section of the reciprocal space, including, apart from the origin, eight points of the reciprocal lattice. The isodiffusion lines obtained have already been published in *Nature* (Born & Begbie 1943), and were discussed there by Dr K. Lonsdale (1943) who found fair agreement with observations.

The case of the close-packed hexagonal lattice is not so favourable, as the number of atomic constants is larger than that of the elastic constants (7:5). Therefore calculation outside the Jahn region is not possible without arbitrary assumptions or other information about these constants. As these calculations would be very involved, only the Jahn case has here been numerically treated for beryl, a crystal for which all the elastic constants are known. Dr Lonsdale has, in a letter to Professor Born, stated 'that photographs of beryl and also of zinc show nothing that does not fit in with the theory quite well'. She adds that in ice and isomorphous NH_4F very strong 'streaks' between Laue spots appear, which, of course, cannot be explained by the Jahn approximation. It will therefore be interesting to work out the complete formulae for these cases, making reasonable assumptions about the constants not determined by the elastic properties. But this work will be postponed for later communication.

1. SUMMARY OF THE FORMULAE FOR THERMAL SCATTERING OF X-RAYS

Assume a beam of X-rays passing through a crystal, and let \mathbf{K} be the wave vector of the incident beam and \mathbf{K}' be the wave vector of the scattered beam. Then, if $\mathbf{K} - \mathbf{K}' = \mathbf{q}$, it can be shown (Born 1943) that the scattering produced by the thermal motion is given by

$$\sigma = \sigma_0 N \sum_{k\alpha} \sum_{k'\beta} S_{\alpha\beta} \left(\frac{\mathbf{q}}{kk'} \right) w_\alpha(k) w_\beta^*(k'), \quad (1.1)$$

where $S_{\alpha\beta} \left(\frac{\mathbf{q}}{kk'} \right)$ are the elements of a matrix, the 'scattering matrix' which in first approximation is reciprocal to the dynamical matrix,

$$S(\mathbf{q}) = kTD(\mathbf{q})^{-1}; \quad (1.2)$$

here k is Boltzmann's constant† and T is the absolute temperature.

The other symbols have the following meanings:

$$\mathbf{w}(k) = \frac{f_k}{\sqrt{m_k}} e^{2\pi i \mathbf{q} \cdot \mathbf{r}(k)} \mathbf{q}; \quad (1.3)$$

$$f_k = \text{modified scattering factor dependent on } T; \quad (1.4)$$

$$N = \text{number of cells in the crystal}; \quad (1.5)$$

$$\sigma_0 = \left(\frac{e^2}{mc^2 r} \right)^2 \frac{1}{2} (1 + \cos^2 \chi) \quad \text{Thomson Factor}; \quad (1.6)$$

$$\left. \begin{aligned} e &= \text{electronic charge;} \\ m &= \text{mass of electron;} \\ c &= \text{velocity of light;} \\ r &= \text{distance of the small scattering volume} \\ &\quad \text{from the point of observation;} \\ \chi &= \text{angle of deflexion of X-ray beam.} \end{aligned} \right\} \quad (1.7)$$

† The fact that k is also used as a suffix elsewhere is not likely to lead to confusion.

2. FACE-CENTRED CUBIC LATTICE. DETERMINATION OF THE DYNAMICAL MATRIX

In the face-centred cubic lattice it is possible to choose a unit cell which contains only one particle, and so the index k can be dropped from all the symbols in which it occurs. This fact is extensively used later, but for the investigation of the symmetry of the lattice it is more convenient to deal with the cubic cell containing four particles (see figure 1), the length of whose side may be denoted by $2a$. The axes of a Cartesian system are taken to be parallel to the sides of this cubic cell.

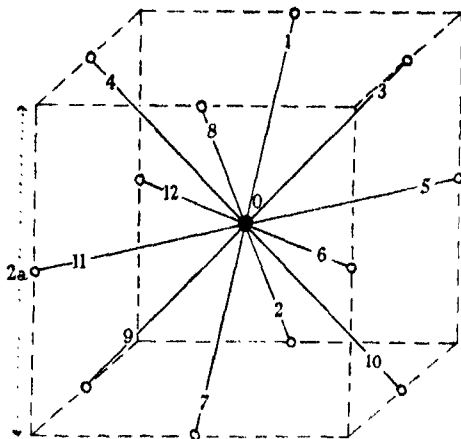


FIGURE 1. First neighbours of a point in the face-centred cubic lattice.

Since in the calculations which follow only the twelve first neighbours of one point, labelled 0 in figure 1, will be taken into account, the number of matrices D^l is $12 + 1 = 13$. The letter l will be used not as the cell index with respect to the smallest cell but as a label attached to the twelve neighbours as figure 1 indicates. The rectangular co-ordinates are given in table 1.

TABLE 1

l	1	2	3	4	5	6	7	8	9	10	11	12
$x_1(l)/a$	0	0	1	-1	1	1	0	0	-1	1	-1	-1
$x_2(l)/a$	1	1	0	0	1	-1	-1	-1	0	0	-1	1
$x_3(l)/a$	1	-1	1	1	0	0	-1	+1	-1	-1	0	0

The complete set of symmetry operations of the lattice, or the space group, may be generated by the following three primitive operations:

A₁. A three-fold axis of rotation, $x_1 = x_2 = x_3$ (see Part I, § 3), to which the matrix

$$T_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (2.1)$$

is attached. In order to use the formula (I (3.9)) the change of the indices l must also be specified. This is given by (I (3.5)) but may be more concisely put in the notation of substitution groups, thus:

$$(1\ 3\ 5)\ (7\ 9\ 11)\ (2\ 4\ 6)\ (8\ 10\ 12),$$

which means that a matrix with a certain number in a cycle is to be derived from the matrix with the number following it in the cycle by the operation $T \dots \hat{T}$, e.g. $D^1 = T_l D^3 \hat{T}_l$.

A₂. A plane of reflexion in the position $x_3 = 0$, with the matrix and change of label (cycles with one element are omitted)

$$T_m = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (1\ 2)\ (3\ 10)\ (4\ 9)\ (7\ 8). \quad (2.2)$$

A₃. A plane of reflexion, $x_1 = x_2$, with the matrix and change of label

$$T_d = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1\ 3)\ (2\ 10)\ (4\ 8)\ (6\ 12)\ (7\ 9). \quad (2.3)$$

As seen from figure 1, D^5 is invariant under (2.2); hence one has

$$\begin{aligned} \begin{pmatrix} D_{11}^5 & D_{12}^5 & D_{13}^5 \\ D_{21}^5 & D_{22}^5 & D_{23}^5 \\ D_{31}^5 & D_{32}^5 & D_{33}^5 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} D_{11}^5 & D_{12}^5 & D_{13}^5 \\ D_{21}^5 & D_{22}^5 & D_{23}^5 \\ D_{31}^5 & D_{32}^5 & D_{33}^5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} D_{11}^5 & D_{12}^5 & -D_{13}^5 \\ D_{21}^5 & D_{22}^5 & -D_{23}^5 \\ -D_{31}^5 & -D_{32}^5 & D_{33}^5 \end{pmatrix}. \end{aligned}$$

From this equation one obtains the identities

$$D_{23}^5 = D_{32}^5 = D_{31}^5 = D_{13}^5 = 0.$$

Since D^5 is invariant also under (2.3), one finds in the same way

$$D_{12}^5 = D_{21}^5, \quad D_{11}^5 = D_{22}^5.$$

Let now

$$D^5 = -\frac{1}{m} \begin{pmatrix} \beta & \gamma & 0 \\ \gamma & \beta & 0 \\ 0 & 0 & \alpha \end{pmatrix}.$$

Then from (2.1) $D^1 = -\frac{1}{m} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & \gamma \\ 0 & \gamma & \beta \end{pmatrix}, \quad D^3 = -\frac{1}{m} \begin{pmatrix} \beta & 0 & \gamma \\ 0 & \alpha & 0 \\ \gamma & 0 & \beta \end{pmatrix}.$

The rest of the D -matrices follow very simply by continued application of the symmetry operations. The results are collected in table 2.

TABLE 2

$$\begin{aligned}
 \left. \begin{matrix} D^1 \\ D^7 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & \gamma \\ 0 & \gamma & \beta \end{pmatrix}, & \left. \begin{matrix} D^2 \\ D^8 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & -\gamma \\ 0 & -\gamma & \beta \end{pmatrix}, \\
 \left. \begin{matrix} D^3 \\ D^9 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \beta & 0 & \gamma \\ 0 & \alpha & 0 \\ \gamma & 0 & \beta \end{pmatrix}, & \left. \begin{matrix} D^4 \\ D^{10} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \beta & 0 & -\gamma \\ 0 & \alpha & 0 \\ -\gamma & 0 & \beta \end{pmatrix}, \\
 \left. \begin{matrix} D^5 \\ D^{11} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \beta & \gamma & 0 \\ \gamma & \beta & 0 \\ 0 & 0 & \alpha \end{pmatrix}, & \left. \begin{matrix} D^6 \\ D^{12} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \beta & -\gamma & 0 \\ -\gamma & \beta & 0 \\ 0 & 0 & \alpha \end{pmatrix}, \\
 D^0 &= -\sum_{l=1}^{12} D^l = \frac{4}{m} (\alpha + 2\beta) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
 \end{aligned}$$

The value of D^0 is obtained from (I (1.6)).

In the case of a Bravais lattice such as this the difference between the matrices $C(\mathbf{q})$ and $D(\mathbf{q})$ disappears. Further, on account of (I (1.6)) one has

$$C^{(0)}(\mathbf{q}) = 0, \quad (2.4)$$

and since each lattice point is a symmetry centre,

$$C^{(1)}(\mathbf{q}) = 0. \quad (2.5)$$

Also, by (I (2.6)),

$$C^{(2)}(\mathbf{q}) = \frac{8\pi^2 a^2}{m} \begin{pmatrix} 2\beta q_1^2 + (\alpha + \beta)(q_2^2 + q_3^2) & 2\gamma q_1 q_2 & 2\gamma q_1 q_3 \\ 2\gamma q_2 q_1 & 2\beta q_2^2 + (\alpha + \beta)(q_3^2 + q_1^2) & 2\gamma q_2 q_3 \\ 2\gamma q_3 q_1 & 2\gamma q_3 q_2 & 2\beta q_3^2 + (\alpha + \beta)(q_1^2 + q_2^2) \end{pmatrix}. \quad (2.6)$$

For this lattice an equation corresponding to (I (2.31)) does not exist because of (2.4) and (2.5), and the equation corresponding to (2.32) is

$$\rho \omega^2 U_\alpha = \sum_{\beta} D'_{\alpha\beta}(\mathbf{q}) U_\beta, \quad (2.7)$$

where $D'(\mathbf{q}) = \rho C^{(2)}(\mathbf{q})$; its elements are given by

$$\begin{pmatrix} D'_{11}(\mathbf{q}) \\ D'_{22}(\mathbf{q}) \\ D'_{33}(\mathbf{q}) \\ D'_{23}(\mathbf{q}) \\ D'_{31}(\mathbf{q}) \\ D'_{12}(\mathbf{q}) \end{pmatrix} = \rho \frac{8\pi^2 a^2}{m} \begin{pmatrix} 2\beta & \alpha + \beta & \alpha + \beta & 0 & 0 & 0 \\ \alpha + \beta & 2\beta & \alpha + \beta & 0 & 0 & 0 \\ \alpha + \beta & \alpha + \beta & 2\beta & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} q_1^2 \\ q_2^2 \\ q_3^2 \\ 2q_2 q_3 \\ 2q_3 q_1 \\ 2q_1 q_2 \end{pmatrix}. \quad (2.8)$$

When this is compared with (I (2.30)) and it is noted that for this lattice the density ρ is $m/2a^3$, the following relations are seen to hold:

$$\left. \begin{aligned} c_{11} = c_{22} = c_{33} &= \frac{4\pi^2}{a} 2\beta, & \text{whence } \alpha &= \frac{a}{8\pi^2} (2c_{44} - c_{11}), \\ c_{44} = c_{55} = c_{66} &= \frac{4\pi^2}{a} (\alpha + \beta), & \beta &= \frac{a}{8\pi^2} c_{11}, \\ c_{23} + c_{44} = c_{31} + c_{55} = c_{12} + c_{66} &= \frac{4\pi^2}{a} 2\gamma, & \gamma &= \frac{a}{8\pi^2} (c_{44} + c_{12}). \end{aligned} \right\} \quad (2.9)$$

The remaining elastic constants all vanish.

The dynamical matrix, defined by (I (1.19)), may now be written as

$$D(\mathbf{q}) = \frac{a}{2\pi^2 m} \begin{pmatrix} F_1 + Cs_1^2 & Cs_1s_2 & Cs_1s_3 \\ Cs_2s_1 & F_2 + Cs_2^2 & Cs_2s_3 \\ Cs_3s_1 & Cs_3s_2 & F_3 + Cs_3^2 \end{pmatrix}, \quad C = c_{44} + c_{12}, \quad (2.10)$$

where

$$\left. \begin{aligned} F_1 &= (2c_{44} - c_{11})(1 - c_2c_3) + c_{11}(2 - c_3c_1 - c_1c_2) - Cs_1^2, \\ F_2 &= (2c_{44} - c_{11})(1 - c_3c_1) + c_{11}(2 - c_1c_2 - c_2c_3) - Cs_2^2, \\ F_3 &= (2c_{44} - c_{11})(1 - c_1c_2) + c_{11}(2 - c_2c_3 - c_3c_1) - Cs_3^2, \end{aligned} \right\} \quad (2.11)$$

with

$$c_\alpha = \cos(2\pi a q_\alpha), \quad s_\alpha = \sin(2\pi a q_\alpha) \quad (\alpha = 1, 2, 3). \quad (2.12)$$

Numerical calculation of $D(\mathbf{q})$ may now be made. For this purpose, the symmetry of the crystal and the periodicity of $D(\mathbf{q})$ itself make it sufficient to consider only a portion of the reciprocal space equivalent to one-eighth of the 'period cube', $-\frac{1}{2a} \leq q_\alpha \leq \frac{1}{2a}$. The results for a representative set of points in this region are given in table 3.

The values of the elastic constants were taken to be those of potassium chloride, although its lattice is neither monatomic nor face-centred. This was done with the purpose of comparing the results with those of Iona (1941) who has calculated the dynamical matrix and the vibrations for this substance. Iona does not work with a next-neighbour model but takes into account the real forces in an ionic lattice: the electrostatic interactions obeying the Coulomb law and a short-range repulsive force. The former was calculated by a procedure due to Ewald (1921), while the latter was found by the method of Born & v. Kármán (1912).

$D(\mathbf{q})$ has here been calculated for the same set of points of reciprocal space that was used by Iona, and if the two tables are compared, the following is noticed:

The values of the matrix elements coincide fairly well in the neighbourhood of the reciprocal lattice points up to distances of about one-quarter of the lattice constant, while very great differences occur in the middle between the lattice points.

The reason for this is, of course, that Iona worked with forces of long range while here only next neighbours are taken into account. This will be of little influence for long waves (small wave number, small q) but very essential for waves of the same order of length as the lattice constant.

TABLE 3. THE DYNAMICAL MATRIX

The values of $mD(\mathbf{q})$ are given in dynes cm.⁻¹.

The units of q_1, q_2, q_3 are π/a , where $a = 3.12 \times 10^{-8}$ cm.

q_1	q_2	q_3	$mD_{11}(\mathbf{q})$ $\times 10^{-3}$	$mD_{22}(\mathbf{q})$ $\times 10^{-3}$	$mD_{33}(\mathbf{q})$ $\times 10^{-3}$	$mD_{23}(\mathbf{q})$ $\times 10^{-3}$	$mD_{31}(\mathbf{q})$ $\times 10^{-3}$	$mD_{12}(\mathbf{q})$ $\times 10^{-3}$
0	0	$\frac{1}{2}$	1.13	1.13	6.97	0.00	0.00	0.00
0	0	$\frac{1}{2}$	4.06	4.06	24.00	0.00	0.00	0.00
0	0	$\frac{3}{2}$	6.94	6.94	41.02	0.00	0.00	0.00
0	0	1	8.12	8.12	47.98	0.00	0.00	0.00
0	$\frac{1}{2}$	$\frac{1}{2}$	2.99	7.18	7.18	2.03	0.00	0.00
0	$\frac{1}{2}$	$\frac{3}{2}$	7.53	7.53	21.68	2.88	0.00	0.00
0	$\frac{1}{2}$	$\frac{3}{2}$	12.09	7.90	36.20	2.03	0.00	0.00
0	$\frac{1}{2}$	1	13.90	8.12	42.20	0.00	0.00	0.00
0	$\frac{1}{2}$	$\frac{1}{2}$	16.06	16.06	16.06	4.06	0.00	0.00
0	$\frac{1}{2}$	$\frac{3}{2}$	24.59	10.43	24.59	2.88	0.00	0.00
0	$\frac{1}{2}$	1	28.05	8.12	28.05	0.00	0.00	0.00
0	$\frac{3}{2}$	$\frac{3}{2}$	37.05	12.93	12.93	2.03	0.00	0.00
0	$\frac{3}{2}$	1	42.20	13.90	13.90	0.00	0.00	0.00
0	1	1	47.98	8.12	8.12	0.00	0.00	0.00
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	8.03	8.03	8.03	2.03	2.03	2.03
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	10.06	10.06	20.03	2.88	2.88	2.03
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	12.09	12.09	32.02	2.03	2.03	2.03
$\frac{1}{2}$	$\frac{1}{2}$	1	12.93	12.93	37.05	0.00	0.00	2.03
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	16.06	16.06	16.06	4.06	2.88	2.88
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	22.06	12.09	22.06	2.88	2.03	2.88
$\frac{1}{2}$	$\frac{1}{2}$	1	24.59	10.43	24.59	0.00	0.00	2.88
$\frac{1}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	32.02	12.09	12.09	2.03	2.03	2.03
$\frac{1}{2}$	$\frac{3}{2}$	1	36.20	7.90	12.09	0.00	0.00	2.03
$\frac{1}{2}$	1	1	41.02	6.94	6.94	0.00	0.00	0.00
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	16.06	16.06	16.06	4.06	4.06	4.06
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	16.06	16.06	16.06	2.88	2.88	4.06
$\frac{1}{2}$	$\frac{1}{2}$	1	16.06	16.06	16.06	0.00	0.00	4.06
$\frac{1}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	20.03	10.06	10.06	2.03	2.88	2.88
$\frac{1}{2}$	$\frac{3}{2}$	1	21.68	7.53	7.53	0.00	0.00	2.88
$\frac{1}{2}$	1	1	24.00	4.06	4.06	0.00	0.00	0.00
$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	8.03	8.03	8.03	2.03	2.03	2.03
$\frac{3}{2}$	$\frac{3}{2}$	1	7.18	7.18	2.99	0.00	0.00	2.03
$\frac{3}{2}$	1	1	6.94	1.19	1.19	0.00	0.00	0.00

The reason that the scattering calculation has been made with a model which certainly is not a good picture of reality is that the method of next neighbours can be generalized; it can be used for all kinds of substances and all kinds of symmetry, while Iona's method is restricted to ionic crystals, the only case where the force law is well known.

3. FACE-CENTRED CUBIC LATTICE. SCATTERING MATRIX AND ISODIFFUSION LINES

The next step is to evaluate the reciprocal matrix. For the general case its elements are obtained at once from (2.10) and are

$$\left. \begin{aligned} D_{11}^{-1} &= \left(\frac{a}{2\pi^2 m} \right)^2 \frac{1}{|D|} [F_2 F_3 + C(s_2^2 F_3 + s_3^2 F_2)], \\ \dots\dots\dots \\ D_{23}^{-1} &= - \left(\frac{a}{2\pi^2 m} \right)^2 \frac{1}{|D|} C s_2 s_3 F_1, \\ \dots\dots\dots \\ |D| &= \left(\frac{a}{2\pi^2 m} \right)^3 [F_1 F_2 F_3 + C(s_1^2 F_2 F_3 + s_2^2 F_3 F_1 + s_3^2 F_1 F_2)]. \end{aligned} \right\} (C = c_{44} + c_{12}) \quad (3.1)$$

In the calculation of the scattering power attention will be confined to the plane $q_3 = 0$. It is therefore sufficient, in order to use the formula (1.1), to know the three reciprocal-matrix elements, $D_{11}^{-1}(\mathbf{q})$, $D_{22}^{-1}(\mathbf{q})$, $D_{12}^{-1}(\mathbf{q})$, which now take the simple form

$$\left. \begin{aligned} D_{11}^{-1}(\mathbf{q}) &= \frac{2\pi^2 m}{a} \frac{F_2 + C s_2^2}{F_1 F_2 + C(F_1 s_2^2 + F_2 s_1^2)}, \\ D_{22}^{-1}(\mathbf{q}) &= \frac{2\pi^2 m}{a} \frac{F_1 + C s_1^2}{F_1 F_2 + C(F_1 s_2^2 + F_2 s_1^2)}, \\ D_{12}^{-1}(\mathbf{q}) &= - \frac{2\pi^2 m}{a} \frac{C s_1 s_2}{F_1 F_2 + C(F_1 s_2^2 + F_2 s_1^2)}. \end{aligned} \right\} \quad (3.2)$$

The values of these have been calculated for a set of points in the plane $q_3 = 0$, twice as close together as those used in the calculation of $D(\mathbf{q})$ in table 3, which was, however, not confined to one plane.

If $\mathbf{q}^{(h)}$ is the vector of a reciprocal lattice point (indices h_1, h_2, h_3) and

$$\bar{\mathbf{q}} = \mathbf{q} - \mathbf{q}^{(h)}, \quad (3.3)$$

then

$$D(\mathbf{q}) = D(\bar{\mathbf{q}}), \quad (3.4)$$

because of the periodicity; it suffices therefore to calculate $D(\mathbf{q})$ for the points $\bar{\mathbf{q}}$ of one single cell of the reciprocal space.

The formula (1.1) for the scattering power of this lattice may be written in the form

$$\sigma = \sigma_0 N k T \frac{|f|^2}{m} d(\mathbf{q}), \quad (3.5)$$

where $|f|^2$ depends exponentially on T (see Born 1942-3) and

$$d(\mathbf{q}) = \sum_{\alpha\beta} D_{\alpha\beta}^{-1}(\bar{\mathbf{q}}) q_\alpha q_\beta. \quad (3.6)$$

The values of the 'diffusion function' (3.6) have been calculated in a region of the plane bounded by $q_1 = q_2 = -1/2a$, $q_1 = q_2 = 5/2a$, so that the nine Laue points whose co-ordinates relative to the axes $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are (0, 0, 0), (1, 0, 0), (0, 1, 0),

$(1, 1, 0)$, $(2, 0, 0)$, $(0, 2, 0)$, $(2, 1, 0)$, $(1, 2, 0)$, $(2, 2, 0)$ are included. It is now possible to plot $d(\mathbf{q})$ against \mathbf{q} along the lines $q_1 = n_1/16a$, $q_2 = n_2/16a$ ($n_1, n_2 = 1, 2, 3, \dots, 8$). From these graphs a set of points on any assigned curve of isodiffusion could be found. Finally, by combining these data, the curves of isodiffusion were drawn for a suitable sequence of values of $d(\mathbf{q})$ and are shown in figure 2.

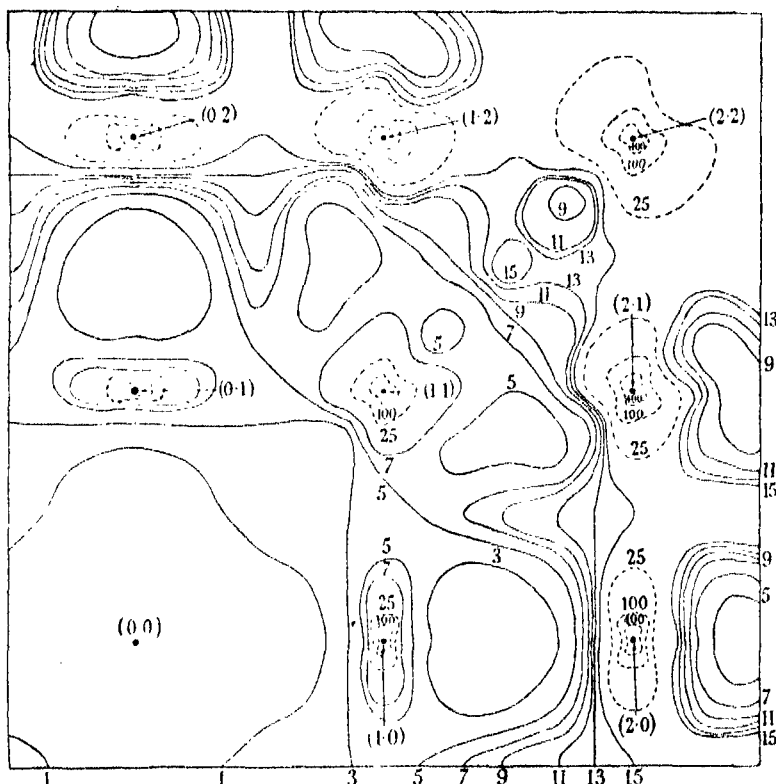


FIGURE 2. Curves of isodiffusion in the plane $q_3 = 0$, for potassium chloride. The values of the elastic constants (Försterling 1920) are $c_{11} = 3.84$, $c_{12} = 0.65$, $c_{44} = 0.04 \times 10^{11}$ dynes/cm.

In the neighbourhood of the Laue points these contours approach closely to those found by Weigle & Smith (1942). But the really interesting features are the 'streaks' or 'bridges' between Laue points. One set of lines form almost squares which divide up the region into unconnected sections. Another set consists of lines diagonal to the axes. With regard to the comparison of these results with experimental data, reference can be made to the note by Dr Lonsdale (1943).

4. HEXAGONAL CLOSE-PACKED LATTICE. THE DYNAMICAL MATRIX

In this lattice there are two particles of equal mass m in the unit cell, distinguished by $k = 1, 2$. The cell vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , can be chosen in such a way that the angle between \mathbf{a}_1 and \mathbf{a}_2 is $2\pi/3$, and \mathbf{a}_3 is at right angles to the plane of \mathbf{a}_1 and \mathbf{a}_2 , in the

direction that makes the system of axes right-handed. Also $|\mathbf{a}_1| = |\mathbf{a}_2| = a$, $|\mathbf{a}_3| = c$.†

Only those matrices $D\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$ will be considered which arise from the first neighbours of the two particles in a chosen unit cell. These two particles are labelled O and O' (black dots, figures 3 and 4). Each point has twelve first neighbours, six in a plane containing the point, three above and three below this plane. The line joining O or O' to a neighbour point $\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$ is labelled by the letter P , and D^P is written for $D\left(\begin{smallmatrix} l \\ kk' \end{smallmatrix}\right)$.

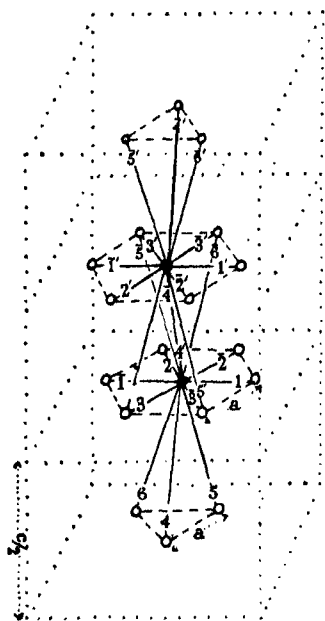


FIGURE 3. First neighbours of the two points in the unit cell of the close-packed hexagonal lattice. The vertical scale has been exaggerated for clarity.

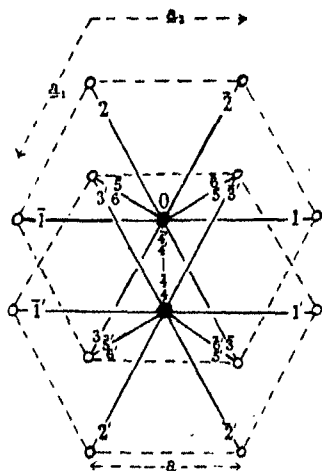


FIGURE 4. Projection of the structure in figure 3 on to a plane perpendicular to the lattice vector \mathbf{a}_3 .

The twelve neighbours of O are again split into two groups p, \bar{p} , those of O' into p', \bar{p}' , as indicated by the following scheme:

$$P \rightarrow \left\{ \begin{array}{l} p \rightarrow 1, 2, 3, 4, 5, 6 \\ \bar{p} \rightarrow \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6} \end{array} \right\} \text{ Neighbours of } O,$$

$$\left\{ \begin{array}{l} p' \rightarrow 1', 2', 3', 4', 5', 6' \\ \bar{p}' \rightarrow \bar{1}', \bar{2}', \bar{3}', \bar{4}', \bar{5}', \bar{6}' \end{array} \right\} \text{ Neighbours of } O'.$$

† The ratio c/a shall remain arbitrary although this does not strictly conform with the expression 'closed packed' (namely: rigid spheres); this special case where $c/a = \sqrt{8/3}$ appears for central forces between next neighbours (Born 1942) and the six constants of the general theory then reduce to a single one.

The six particles that lie in a plane containing the point O are taken to be of the type $k = 1$, the six that lie in the plane containing O' are of type $k = 2$, and so on alternately, so that the particles of one type lie on a set of planes that interleave the set of planes which contain the other type of particles.

The connexion between the symbols P and $\begin{pmatrix} l \\ kk' \end{pmatrix}$ is given in table 4.

TABLE 4

P	1	2	3	4	5	6	1	2	3	4	5	6
l_1	0	-1	1	0	-1	-1	0	-1	1	0	-1	-1
l_2	1	-1	0	0	0	-1	-1	0	1	0	-1	0
l_3	0	0	0	-1	-1	-1	0	0	0	0	0	0
(kk')	(11)	(11)	(11)	(21)	(21)	(21)	(11)	(11)	(11)	(21)	(21)	(21)
P	1'	2'	3'	4'	5'	6'	1'	2'	3'	4'	5'	6'
l_1	0	1	-1	0	1	1	0	1	-1	0	1	1
l_2	1	0	-1	0	1	0	-1	1	0	0	0	1
l_3	0	0	0	0	0	0	0	0	0	1	1	1
(kk')	(22)	(22)	(22)	(12)	(12)	(12)	(22)	(22)	(22)	(12)	(12)	(12)

The Cartesian co-ordinate system S is so oriented that x_1 lies along the positive direction of \mathbf{a}_2 and x_3 along the positive direction of \mathbf{a}_3 .

The group of symmetry operations of the lattice can be generated by the following three primitive operations.

A₁. A six-fold axis of rotatory inversion in the direction of the x_3 -axis, that is to say, a rotation of amount $\frac{1}{3}\pi$ about the x_3 -axis together with an inversion in the origin. This operation can be represented by the matrix

$$T_r = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (4.1)$$

The change of label P is specified, in an analogous way to that used for the face-centred cubic lattice, by the substitution group

$$(3 \ 2 \ 1) \ (\bar{1} \ \bar{2} \ \bar{3}) \ (5 \ \bar{4} \ 6 \ \bar{6} \ 4 \ \bar{5}).$$

A₂. A two-fold axis of rotation, namely, the x_3 -axis. This has the matrix and change of label,

$$T_u = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (p\bar{p}). \quad (4.2)$$

A₃. A glide plane. This involves a reflexion in the plane $x_2 = -a/\sqrt{3}$ followed by a translation of amount $x_1 = 0$, $x_2 = 0$, $x_3 = \frac{1}{2}c$ parallel to this plane. This is equivalent to a reflexion in $x_2 = 0$ together with a translation of amount

$$x_1 = 0, \quad x_2 = -2a/\sqrt{3}, \quad x_3 = \frac{1}{2}c.$$

The latter only affects the change of label, and the T -matrix is therefore that due to the reflexion in $x_2 = 0$, i.e.

$$T_g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = -T_u, \quad (pp')(\bar{p}\bar{p}'). \quad (4.3)$$

Use will be made of the following three combinations of these elementary operations:

A₄. A reflexion in the plane $x_3 = 0$. This is represented by the matrix

$$T_r^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (5\bar{6}) \quad (6\bar{5}) \quad (4\bar{4}). \quad (4.4)$$

A₅. A reflexion in the plane $x_1 = 0$:

$$T_u T_r^3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1\bar{1}) \quad (2\bar{2}) \quad (3\bar{3}) \quad (56) \quad (\bar{5}\bar{6}). \quad (4.5)$$

A₆. From **A₂** it is found that

$$D^p = T_u D^p \tilde{T}_u = T_g D^p \tilde{T}_g = D^{p'}. \quad (4.6)$$

Similarly

$$D^p = D^{p'}.$$

A₇. Finally, the general relation (I (1.4)) leads here to the identities

$$\left. \begin{aligned} D^1 &= \tilde{D}^1, & D^4 &= \tilde{D}^4, & D^{1'} &= \tilde{D}^{1'}, & D^{4'} &= \tilde{D}^{4'}, \\ D^2 &= \tilde{D}^2, & D^5 &= \tilde{D}^5, & D^{2'} &= \tilde{D}^{2'}, & D^{5'} &= \tilde{D}^{5'}, \\ D^3 &= \tilde{D}^3, & D^6 &= \tilde{D}^6, & D^{3'} &= \tilde{D}^{3'}, & D^{6'} &= \tilde{D}^{6'}. \end{aligned} \right\} \quad (4.7)$$

Only the twelve matrices D^p and $D^{p'}$ need be calculated, since **A₆** gives the remaining ones at once. From (4.4)

$$\begin{pmatrix} D_{11}^1 & D_{12}^1 & D_{13}^1 \\ D_{21}^1 & D_{22}^1 & D_{23}^1 \\ D_{31}^1 & D_{32}^1 & D_{33}^1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} D_{11}^1 & D_{12}^1 & D_{13}^1 \\ D_{21}^1 & D_{22}^1 & D_{23}^1 \\ D_{31}^1 & D_{32}^1 & D_{33}^1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} D_{11}^1 & D_{12}^1 & -D_{13}^1 \\ D_{21}^1 & D_{22}^1 & -D_{23}^1 \\ -D_{31}^1 & -D_{32}^1 & D_{33}^1 \end{pmatrix}$$

Hence

$$D_{13}^1 = D_{31}^1 = D_{23}^1 = D_{32}^1 = 0.$$

Let†

$$D^1 = -\frac{1}{m} \begin{pmatrix} \alpha + 2\beta & \delta & 0 \\ \epsilon & \alpha - 2\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}.$$

Now by (4.2)

$$D^{\bar{1}} = -\frac{1}{m} \begin{pmatrix} \alpha + 2\beta & -\delta & 0 \\ -\epsilon & \alpha - 2\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}.$$

But by (4.7)

$$D^{\bar{1}} = -\frac{1}{m} \begin{pmatrix} \alpha + 2\beta & \epsilon & 0 \\ \delta & \alpha - 2\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}.$$

Hence

$$\epsilon = -\delta.$$

Now D^2 , $D^{\bar{2}}$ and D^3 , $D^{\bar{3}}$ may be calculated directly from (4.1); thus $D^2 = T_r D^1 \bar{T}_r$ and $D^3 = T_r D^2 \bar{T}_r$, where for a general D -matrix,

$$T_r D \bar{T}_r = \begin{pmatrix} \frac{1}{4}(D_{11} + \sqrt{3}D_{21} + \sqrt{3}D_{12} + 3D_{22}) & \frac{1}{4}(-\sqrt{3}D_{11} - 3D_{21} + D_{12} + \sqrt{3}D_{22}) & \frac{1}{2}(D_{13} + \sqrt{3}D_{23}) \\ \frac{1}{4}(D_{11} + D_{21} - 3D_{12} + \sqrt{3}D_{23}) & \frac{1}{4}(3D_{11} - \sqrt{3}D_{21} - \sqrt{3}D_{12} + D_{22}) & \frac{1}{2}(-\sqrt{3}D_{13} + D_{23}) \\ \frac{1}{2}(D_{31} + \sqrt{3}D_{32}) & \frac{1}{2}(-\sqrt{3}D_{31} + D_{32}) & D_{33} \end{pmatrix}.$$

Again $D^4 = D^{\bar{4}}$ by (4.6). But $D^4 = \bar{D}^4$ by (4.7). Hence $D^4 = \bar{D}^4$. Moreover, from (4.5)

$$\begin{pmatrix} D_{11}^4 & D_{12}^4 & D_{13}^4 \\ D_{21}^4 & D_{22}^4 & D_{23}^4 \\ D_{31}^4 & D_{32}^4 & D_{33}^4 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} D_{11}^4 & D_{12}^4 & D_{13}^4 \\ D_{21}^4 & D_{22}^4 & D_{23}^4 \\ D_{31}^4 & D_{32}^4 & D_{33}^4 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

from which

$$D_{12}^4 = D_{21}^4 = D_{13}^4 = D_{31}^4 = 0.$$

Let‡

$$D^4 = -\frac{1}{m} \begin{pmatrix} \lambda + 2\mu & 0 & 0 \\ 0 & \lambda - 2\mu & 2\sigma \\ 0 & 2\sigma & \nu \end{pmatrix}.$$

Then $D^{\bar{3}}$, D^6 , $D^{\bar{4}}$, D^5 , $D^{\bar{5}}$ can be calculated from (4.1) with the help of the formula for $T_r D \bar{T}_r$ above.

The final results are collected in table 5.

† The notation is chosen to avoid fractions.

‡ The notation is again chosen to avoid fractions.

TABLE 5

$$\begin{aligned}
\left. \begin{matrix} D^1 \\ D^{\bar{1}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha+2\beta & \delta & 0 \\ -\delta & \alpha-2\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, & \left. \begin{matrix} D^{\bar{1}} \\ D^1 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha+2\beta & -\delta & 0 \\ \delta & \alpha-2\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \\
\left. \begin{matrix} D^2 \\ D^{\bar{2}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha-\beta & \delta-\sqrt{3}\beta & 0 \\ -\delta-\sqrt{3}\beta & \alpha+\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, & \left. \begin{matrix} D^{\bar{2}} \\ D^2 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha-\beta & -\delta+\sqrt{3}\beta & 0 \\ \delta+\sqrt{3}\beta & \alpha+\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \\
\left. \begin{matrix} D^3 \\ D^{\bar{3}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha-\beta & \delta+\sqrt{3}\beta & 0 \\ -\delta+\sqrt{3}\beta & \alpha+\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, & \left. \begin{matrix} D^{\bar{3}} \\ D^3 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \alpha-\beta & -\delta-\sqrt{3}\beta & 0 \\ \delta-\sqrt{3}\beta & \alpha+\beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \\
\left. \begin{matrix} D^4 \\ D^{\bar{4}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda+2\mu & 0 & 0 \\ 0 & \lambda-2\mu & 2\sigma \\ 0 & 2\sigma & \nu \end{pmatrix}, & \left. \begin{matrix} D^4 \\ D^{\bar{4}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda+2\mu & 0 & 0 \\ 0 & \lambda-2\mu & -2\sigma \\ 0 & -2\sigma & \nu \end{pmatrix}, \\
\left. \begin{matrix} D^5 \\ D^{\bar{5}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda-\mu & -\sqrt{3}\mu & -\sqrt{3}\sigma \\ -\sqrt{3}\mu & \lambda+\mu & -\sigma \\ -\sqrt{3}\sigma & -\sigma & \nu \end{pmatrix}, & \left. \begin{matrix} D^{\bar{5}} \\ D^5 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda-\mu & \sqrt{3}\mu & -\sqrt{3}\sigma \\ \sqrt{3}\mu & \lambda-\mu & \sigma \\ -\sqrt{3}\sigma & \sigma & \nu \end{pmatrix}, \\
\left. \begin{matrix} D^6 \\ D^{\bar{6}} \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda-\mu & \sqrt{3}\mu & \sqrt{3}\sigma \\ \sqrt{3}\mu & \lambda+\mu & -\sigma \\ \sqrt{3}\sigma & -\sigma & \nu \end{pmatrix}, & \left. \begin{matrix} D^{\bar{6}} \\ D^6 \end{matrix} \right\} &= -\frac{1}{m} \begin{pmatrix} \lambda-\mu & -\sqrt{3}\mu & \sqrt{3}\sigma \\ -\sqrt{3}\mu & \lambda+\mu & \sigma \\ \sqrt{3}\sigma & \sigma & \nu \end{pmatrix}, \\
\left. \begin{matrix} D^0 \\ D^{0'} \end{matrix} \right\} &= \frac{1}{m} \begin{pmatrix} 6\alpha+6\lambda & 0 & 0 \\ 0 & 6\alpha+6\lambda & 0 \\ 0 & 0 & 6\gamma+6\nu \end{pmatrix}.
\end{aligned}$$

From (4.6) then

$$D^p = D^{\bar{p}'}, \quad D^{p'} = D^{\bar{p}}, \quad (4.8)$$

while

$$\mathbf{r}^p = -\mathbf{r}^{p'}, \quad \mathbf{r}^{p'} = -\mathbf{r}^p. \quad (4.9)$$

Therefore it follows from the construction of $C\left(\begin{smallmatrix} \mathbf{q} \\ kk' \end{smallmatrix}\right)$, given in (I (1.25)) that

$$C\left(\begin{smallmatrix} \mathbf{q} \\ 22 \end{smallmatrix}\right) = C\left(\begin{smallmatrix} -\mathbf{q} \\ 11 \end{smallmatrix}\right), \quad C\left(\begin{smallmatrix} \mathbf{q} \\ 12 \end{smallmatrix}\right) = C\left(\begin{smallmatrix} -\mathbf{q} \\ 21 \end{smallmatrix}\right). \quad (4.10)$$

So it is only necessary to calculate the terms in the expansions of

$$C\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) = \sum_{P=0, \substack{1, 2, 3 \\ \bar{1}, \bar{2}, \bar{3}}} D^P e^{-2\pi i(\mathbf{q}, \mathbf{r}^P)} \quad (4.11)$$

and

$$C\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) = \sum_{P=\substack{4, 5, 6 \\ \bar{4}, \bar{5}, \bar{6}}} D^P e^{-2\pi i(\mathbf{q}, \mathbf{r}^P)}. \quad (4.12)$$

The values of \mathbf{r}^P are tabulated in table 6.

TABLE 6

P	1	2	3	1	2	3	P	4	5	6	4	5	6
$\frac{1}{a}x_1\left(\begin{smallmatrix} l \\ 11 \end{smallmatrix}\right)$	1	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{a}x_1\left(\begin{smallmatrix} l \\ 21 \end{smallmatrix}\right)$	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{a}x_2\left(\begin{smallmatrix} l \\ 11 \end{smallmatrix}\right)$	0	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	0	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	$\frac{1}{a}x_2\left(\begin{smallmatrix} l \\ 21 \end{smallmatrix}\right)$	$-\frac{1}{\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$-\frac{1}{\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$
$\frac{1}{c}x_3\left(\begin{smallmatrix} l \\ 11 \end{smallmatrix}\right)$	0	0	0	0	0	0	$\frac{1}{c}x_3\left(\begin{smallmatrix} l \\ 21 \end{smallmatrix}\right)$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

The first three terms in the expansions of $C\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right)$ and $C\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right)$ now become

$$\left. \begin{aligned} C^{(0)}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} \begin{pmatrix} 6\lambda & 0 & 0 \\ 0 & 6\lambda & 0 \\ 0 & 0 & 6\nu \end{pmatrix}, & C^{(0)}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= -\frac{1}{m} \begin{pmatrix} 6\lambda & 0 & 0 \\ 0 & 6\lambda & 0 \\ 0 & 0 & 6\nu \end{pmatrix}; \\ \\ C^{(1)}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= 0, & C^{(1)}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= -\frac{2\pi i a 2\sqrt{3}\mu}{m} \begin{pmatrix} q_2 & q_1 & 0 \\ q_1 & -q_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \\ \\ C^{(2)}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= 4\pi^2 \frac{3a^2}{2m} \begin{pmatrix} (\alpha + \beta)q_1^2 + (\alpha - \beta)q_2^2 & 2\beta q_1 q_2 & 0 \\ 2\beta q_1 q_2 & (\alpha - \beta)q_1^2 + (\alpha + \beta)q_2^2 & 0 \\ 0 & 0 & \gamma(q_1^2 + q_2^2) \end{pmatrix}, \\ \\ C^{(2)}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= \frac{2\pi^2}{m} \begin{pmatrix} a^2[(\lambda - \mu)q_1^2 + (\lambda + \mu)q_2^2] + \frac{3}{2}c^2q_3^2 & -2a^2\mu q_1 q_2 & 2\sqrt{3}ac\sigma q_1 q_2 \\ -2a^2\mu q_2 q_1 & a^2[(\lambda + \mu)q_1^2 + (\lambda - \mu)q_2^2] + \frac{3}{2}c^2q_3^2 & 2\sqrt{3}ac\sigma q_2 q_3 \\ 2\sqrt{3}ac\sigma q_3 q_1 & 2\sqrt{3}ac\sigma q_3 q_2 & \nu[a^2(q_1^2 + q_2^2) + \frac{3}{2}c^2q_3^2] \end{pmatrix}. \end{aligned} \right\} \quad (4.13)$$

If these values are substituted into (I (2.31)) then

$$\frac{1}{m} \begin{pmatrix} 6\lambda & 0 & 0 \\ 0 & 6\lambda & 0 \\ 0 & 0 & 6\nu \end{pmatrix} (\mathbf{W}^{(1)}(1) - \mathbf{W}^{(1)}(2)) = -\frac{2\pi i a 2\sqrt{3}\mu}{\sqrt{m}} \begin{pmatrix} q_2 & q_1 & 0 \\ q_1 & -q_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{U}, \quad (4.14)$$

$$\text{whence} \quad \mathbf{W}^{(1)}(1) - \mathbf{W}^{(1)}(2) = -\frac{2\pi i a 2\sqrt{3}\mu\sqrt{m}}{6\lambda} \begin{pmatrix} q_2 & q_1 & 0 \\ q_1 & -q_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{U}. \quad (4.15)$$

If this, along with (4.13), is substituted in (I (2.32)) the result is

$$\rho\omega^2 U_\alpha = \sum_\beta D'_{\alpha\beta}(\mathbf{q}) U_\beta, \quad (4.16)$$

where now

$$\begin{pmatrix} D'_{11}(\mathbf{q}) \\ D'_{22}(\mathbf{q}) \\ D'_{33}(\mathbf{q}) \\ D'_{23}(\mathbf{q}) \\ D'_{31}(\mathbf{q}) \\ D'_{12}(\mathbf{q}) \end{pmatrix} = \frac{4\pi^2 \rho a^3}{m} \begin{pmatrix} -\frac{2\mu^2}{\lambda} + 3\alpha + 3\beta + \lambda - \mu & -\frac{2\mu^2}{\lambda} + 3\alpha - 3\beta + \lambda + \mu & \frac{3c^2}{2a^2}\lambda & 0 & 0 & 0 \\ -\frac{2\mu^2}{\lambda} + 3\alpha - 3\beta + \lambda + \mu & -\frac{2\mu^2}{\lambda} + 3\alpha + 3\beta + \lambda - \mu & \frac{3c^2}{2a^2}\lambda & 0 & 0 & 0 \\ 3\gamma + \nu & 3\gamma + \nu & \frac{3c^2}{2a^2}\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{3}\frac{c}{a}\sigma & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{3}\frac{c}{a}\sigma & 0 \\ 0 & 0 & 0 & 0 & 0 & 3\beta - \mu \end{pmatrix} \begin{pmatrix} q_1^2 \\ q_2^2 \\ q_3^2 \\ 2q_2q_3 \\ 2q_3q_1 \\ 2q_1q_2 \end{pmatrix}. \quad (4.17)$$

By comparing this with (I (2.30)) and noting that for this lattice $\rho = \frac{4m}{\sqrt{3}a^2c}$, one has

$$\left. \begin{aligned} c_{11} = c_{22} &= \frac{16\pi^2}{\sqrt{3}c} \left(-\frac{2\mu^2}{\lambda} + 3\alpha + 3\beta + \lambda - \mu \right), \\ c_{33} &= \frac{16\pi^2 c}{\sqrt{3}a^2} \nu, \\ c_{44} = c_{55} &= \frac{16\pi^2 c}{\sqrt{3}a^2} \frac{3}{2} \lambda, \\ c_{66} &= \frac{16\pi^2}{\sqrt{3}c} \left(-\frac{2\mu^2}{\lambda} + 3\alpha - 3\beta + \lambda + \mu \right), \\ c_{23} = c_{31} &= \frac{16\pi^2}{\sqrt{3}a} \left(2\sqrt{3}\sigma - \frac{3c}{2a}\lambda \right), \\ c_{12} &= \frac{16\pi^2}{\sqrt{3}c} \left(\frac{2\mu^2}{\lambda} - 3\alpha + 9\beta - \lambda - 3\mu \right). \end{aligned} \right\} \quad (4.18)$$

The remaining elastic constants all vanish. It will be noted that the relation characteristic for hexagonal symmetry, $c_{66} = \frac{1}{2}(c_{11} - c_{12})$, is satisfied. The identities (I (1.7)) lead to a further relation between the arbitrary constants used in forming the dynamical matrix, namely,

$$\frac{3}{2}\lambda \frac{c^2}{a^2} = 3\gamma + \nu \quad \text{or} \quad 4\lambda = 3\gamma + \nu. \quad (4.19)$$

By this the number 8 of atomic constants is reduced to 7, while that of the elastic constants is 5. Thence it is not possible to express the atomic constants unambiguously in terms of the elastic constants. Therefore the complete formulae for the dynamical matrix are now given in terms of the atomic constants:

$$\left. \begin{aligned}
 D_{11}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} [6\lambda + 4(\alpha + 2\beta) (1 - \cos^2(\pi a q_1)) + 4(\alpha - \beta) \{1 - \cos(\pi a q_1) \cos(\sqrt{3} \pi a q_2)\}], \\
 D_{22}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} [6\lambda + 4(\alpha - 2\beta) (1 - \cos^2(\pi a q_1)) + 4(\alpha + \beta) \{1 - \cos(\pi a q_1) \cos(\sqrt{3} \pi a q_2)\}], \\
 D_{33}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} [6\nu + 4\gamma \{2 - \cos^2(\pi a q_1) - \cos(\pi a q_1) \cos(\sqrt{3} \pi a q_2)\}], \\
 D_{23}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= D_{32}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) = D_{31}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) = D_{13}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) = 0, \\
 D_{12}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} [4\sqrt{3} \beta \sin(\pi a q_1) \sin(\sqrt{3} \pi a q_2) + 4i\delta \sin(\pi a q_1) \{\cos(\pi a q_1) - \cos(\sqrt{3} \pi a q_2)\}], \\
 D_{21}\left(\begin{smallmatrix} \mathbf{q} \\ 11 \end{smallmatrix}\right) &= \frac{1}{m} [4\sqrt{3} \beta \sin(\pi a q_1) \sin(\sqrt{3} \pi a q_2) - 4i\delta \sin(\pi a q_1) \{\cos(\pi a q_1) - \cos(\sqrt{3} \pi a q_2)\}].
 \end{aligned} \right\} \quad (4.20)$$

$$\left. \begin{aligned}
 D_{11}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= \frac{1}{m} \left[-2(\lambda + 2\mu) \cos(\pi c q_3) \exp\left(\frac{2\pi i a}{\sqrt{3}} q_2\right) - 4(\lambda - \mu) \cos(\pi c q_3) \cos(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right) \right], \\
 D_{22}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= \frac{1}{m} \left[-2(\lambda - 2\mu) \cos(\pi c q_3) \exp\left(\frac{2\pi i a}{\sqrt{3}} q_2\right) - 4(\lambda + \mu) \cos(\pi c q_3) \cos(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right) \right], \\
 D_{33}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= -\frac{1}{m} 2\nu \cos(\pi c q_3) \left[\exp\left(\frac{2\pi i a}{\sqrt{3}} q_2\right) + 2 \cos(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right) \right], \\
 D_{23}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= D_{32}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) = \frac{4i}{m} \sigma \sin(\pi c q_3) \left[\cos(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right) - \exp\left(\frac{2\pi i a}{\sqrt{3}} q_2\right) \right], \\
 D_{31}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= D_{13}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) = \frac{4\sqrt{3}}{m} \sigma \sin(\pi c q_3) \sin(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right), \\
 D_{12}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) &= D_{21}\left(\begin{smallmatrix} \mathbf{q} \\ 21 \end{smallmatrix}\right) = -\frac{4\sqrt{3}}{m} i\mu \cos(\pi c q_3) \sin(\pi a q_1) \exp\left(-\frac{\pi i a}{\sqrt{3}} q_2\right).
 \end{aligned} \right\} \quad (4.21)^\dagger$$

In the Jahn case the elements of the dynamical matrix can be expressed in terms of elastic constants, and for the matrix $D'(\bar{\mathbf{q}})$, defined by (I (2.30)), one obtains

$$\begin{pmatrix} D'_{11}(\bar{\mathbf{q}}) \\ D'_{22}(\bar{\mathbf{q}}) \\ D'_{33}(\bar{\mathbf{q}}) \\ D'_{23}(\bar{\mathbf{q}}) \\ D'_{31}(\bar{\mathbf{q}}) \\ D'_{12}(\bar{\mathbf{q}}) \end{pmatrix} = \begin{pmatrix} c_{11} & \frac{1}{2}(c_{11} - c_{12}) & c_{44} & 0 & 0 & 0 \\ \frac{1}{2}(c_{11} - c_{12}) & c_{11} & c_{44} & 0 & 0 & 0 \\ c_{44} & c_{44} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2}(c_{13} + c_{44}) & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}(c_{13} + c_{44}) & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(c_{11} + c_{12}) \end{pmatrix} \begin{pmatrix} \bar{q}_1^2 \\ \bar{q}_2^2 \\ \bar{q}_3^2 \\ 2\bar{q}_2\bar{q}_3 \\ 2\bar{q}_3\bar{q}_1 \\ 2\bar{q}_1\bar{q}_2 \end{pmatrix}. \quad (4.22)$$

† (Correction in proof.) On the right-hand side of (4.21) the factor $\exp\left[-2\pi i\left(\frac{a q_2}{\sqrt{3}} - \frac{c q_3}{2}\right)\right]$ has to be added.

5. HEXAGONAL CLOSE-PACKED LATTICE. THE SCATTERING MATRIX
AND ISODIFFUSION LINES FOR THE JAHN CASE

For the Jahn approximation one can replace in the diffusion function (3.6) the q_α outside of the periodic function $D'_{\alpha\beta}{}^{-1}$ by their values in the reciprocal lattice point, $q_\alpha^{(h)}$, for the neighbourhood of which the isodiffusion lines are to be drawn:

$$d(\mathbf{q}) = \sum_{\alpha\beta} D'_{\alpha\beta}{}^{-1}(\bar{\mathbf{q}}) q_\alpha^{(h)} q_\beta^{(h)}. \quad (5.1)$$

In order to express the reciprocal matrix in a simple way the following abbreviations are used:

$$\bar{q}_1 = \bar{q}l, \quad \bar{q}_2 = \bar{q}m, \quad \bar{q}_3 = \bar{q}n, \quad l^2 + m^2 + n^2 = 1, \quad (5.2)$$

$$a = c_{11}, \quad b = \frac{1}{2}(c_{11} - c_{12}), \quad c = c_{44}, \quad d = c_{33}, \quad f = c_{13} + c_{44}, \quad (5.3)$$

$$\left. \begin{aligned} A_1 &= al^2 + bm^2 + cn^2, \\ A_2 &= bl^2 + am^2 + cn^2, \\ A_3 &= c(l^2 + m^2) + dn^2. \end{aligned} \right\} \quad (5.4)$$

Then the elements of the reciprocal matrix are given by

$$D'_{\alpha\beta}{}^{-1}(\mathbf{q}) = \frac{\text{adj}_{\alpha\beta} D'(\mathbf{q})}{\det D'(\mathbf{q})}, \quad (5.5)$$

where

$$\begin{aligned} \frac{1}{\bar{q}^3} \det D'(\bar{\mathbf{q}}) &= A_1 A_2 A_3 - A_3(a-b)^2 l^2 m^2 - A_1 f^2 m^2 n^2 - A_2 f^2 n^2 l^2 + 2(a-b) f^2 l^2 m^2 n^2 \\ &= [b + (c-b)n^2] \cdot \{[a + (c-a)n^2]\{c + (d-c)n^2\} - f^2 n^2(1-n^2)\}. \end{aligned} \quad (5.6)$$

$$\left. \begin{aligned} \frac{1}{\bar{q}^4} \text{adj}_{11} D'(\bar{\mathbf{q}}) &= A_2 A_3 - f^2 m^2 n^2 = l^2[(b-a)\{c + (d-c)n^2\} + f^2 n^2] \\ &\quad + [a + (c-a)n^2][c + (d-c)n^2] - f^2 n^2(1-n^2), \\ \frac{1}{\bar{q}^4} \text{adj}_{22} D'(\bar{\mathbf{q}}) &= A_1 A_3 - f^2 n^2 l^2 = m^2[(b-a)\{c + (d-c)n^2\} + f^2 n^2] \\ &\quad + [a + (c-a)n^2][c + (d-c)n^2] - f^2 n^2(1-n^2), \\ \frac{1}{\bar{q}^4} \text{adj}_{33} D'(\bar{\mathbf{q}}) &= A_1 A_2 - (a-b)^2 l^2 m^2 \\ &= [b + (c-b)n^2] \cdot [a + (c-a)n^2], \\ \frac{1}{\bar{q}^4} \text{adj}_{23} D'(\bar{\mathbf{q}}) &= -A_1 f m n + (a-b) f l^2 m n = -f m n [b + (c-b)n^2], \\ \frac{1}{\bar{q}^4} \text{adj}_{31} D'(\bar{\mathbf{q}}) &= -A_2 f n l + (a-b) f l m^2 n = -f n l [b + (c-b)n^2], \\ \frac{1}{\bar{q}^4} \text{adj}_{12} D'(\bar{\mathbf{q}}) &= -A_3(a-b) l m + f^2 l m n^2 = l m [f^2 n^2 - (a-b)\{c + (d-c)n^2\}]. \end{aligned} \right\} \quad (5.7)$$

The case $q_3^{(h)} = 0$ will be considered. Then after elementary calculation

$$d(\mathbf{q}) = D_{11}^{-1} q_1^{(h)2} + D_{22}^{-1} q_2^{(h)2} + 2D_{12}^{-1} q_1^{(h)} q_2^{(h)} \\ = \frac{1}{\bar{q}^2} \left(\frac{\{lq_1^{(h)} + mq_2^{(h)}\}^2 \{(b-a)(c + [d-c]n^2) + f^2n^2\}}{[b + (c-b)n^2][\{a + (c-a)n^2\}\{c + (d-c)n^2\} - f^2n^2(1-n^2)]} + \frac{q_1^{(h)2} + q_2^{(h)2}}{b + (c-b)n^2} \right).$$

If now θ is the angle between $\mathbf{q}^{(h)}$ and $\bar{\mathbf{q}} = \mathbf{q} - \mathbf{q}^{(h)}$, and $q^{(h)}$ the length of the vector $\mathbf{q}^{(h)}$,

$$q^{(h)} \cos \theta = q_1^{(h)} l + q_2^{(h)} m, \quad q^{(h)2} = q_1^{(h)2} + q_2^{(h)2}, \quad (5.8)$$

$$d(\mathbf{q}) = \left(\frac{q^{(h)}}{\bar{q}} \right)^2 \left\{ \frac{1}{b + (c-b)n^2} + \frac{\cos^2 \theta (1-n^2) \{(b-a)[c + (d-c)n^2] + f^2n^2\}}{[b + (c-b)n^2][\{a + (c-a)n^2\}\{c + (d-c)n^2\} - f^2n^2(1-n^2)]} \right\}. \quad (5.9)$$

The isodiffusion surfaces around the lattice point (h) are given by $d(\mathbf{q}) = \text{const.}$ Their cross-sections in the plane $n = 0$ are particularly simple; one has

$$d(\mathbf{q}) = \left(\frac{q^{(h)}}{\bar{q}} \right)^2 \left(\frac{\cos^2 \theta}{a} + \frac{\sin^2 \theta}{b} \right) \quad (n = 0). \quad (5.10)$$

The curves $d(\mathbf{q}) = \text{const.}$ have the same shape for all points $\mathbf{q}^{(h)}$ which lie in the plane $q_3 = 0$, and differ only in size and orientation. A part of this plane is shown in figure 5 and includes the reciprocal lattice points $(0, 0)$, $(0, 1)$, $(-1, 0)$, $(-1, 1)$

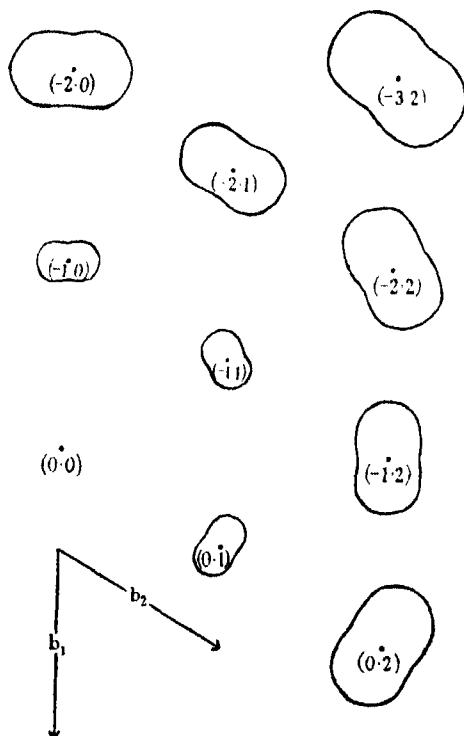


FIGURE 5. Curves of isodiffusion in the plane $q_3 = 0$ for beryl. The values of the elastic constants (Voigt 1887) are $c_{11} = 27.5$, $c_{22} = 24.1$, $c_{12} = 9.80$, $c_{13} = 6.74$, $c_{44} = 6.66 \times 10^{11}$ dynes/cm.

$(-1, 2)$, $(0, 2)$, $(-2, 0)$, $(-2, 1)$, $(-2, 2)$, $(-3, 2)$, each surrounded by an isodiffusion line. The elastic constants are taken as those of beryl (Voigt 1887).

Professor Born suggested this work to me and I am very greatly indebted to him for his advice on numerous occasions. I would also like to thank Dr H. W. Peng for many valuable suggestions.

REFERENCES

- Born, M. 1942 *Proc. Camb. Phil. Soc.* **38**, 82.
 Born, M. 1942-3. *Rep. Progr. Phys.* **9**, 294.
 Born, M. & Begbie, G. H. 1943 *Nature*, **152**, 19.
 Born, M. & v. Kármán, Th. 1912 *Phys. Z.* **13**, 297.
 Born, M. & v. Kármán, Th. 1913 *Phys. Z.* **14**, 15, 65.
 Ewald, P. 1921 *Ann. Phys., Lpz.*, **64**, 253.
 Försterling, K. 1920 *Z. Phys.* **2**, 172.
 Iona, M. 1941 *Phys. Rev.* **60**, 822.
 Jahn, H. A. 1941 *Nature*, **147**, 511.
 Jahn, H. A. 1942a *Proc. Roy. Soc. A*, **179**, 320.
 Jahn, H. A. 1942b *Proc. Roy. Soc. A*, **180**, 476.
 Lonsdale, K. 1943 *Nature*, **152**, 20.
 Voigt, W. 1887 *Wied. Ann. Phys.* **31**, 474.
 Weigle, J. & Smith, C. S. 1942 *Phys. Rev.* **61**, 23.

The aerodynamic drag of the earth's surface and the value of von Karman's constant in the lower atmosphere

By P. A. SHEPPARD, *Imperial College*

(Communicated by D. Brunt, F.R.S.—Received 9 February 1946)

[Plates 1, 2]

The drag which the earth's surface exerts on the wind has been measured directly by observing the deflexion of a horizontal plate, floating in oil and under torsional control, and forming part of the earth's surface.

Simultaneous measurements of the velocity profile above the surface have enabled von Karman's constant for the lower atmosphere to be deduced. The value obtained under moderately unstable atmospheric stratification is 0.46, which is considered to be in good agreement with Nikuradse's value of 0.40, obtained from the flow of water through pipes under isothermal conditions. Evidence is also adduced for a considerable variation of von Karman's constant with atmospheric stability.

The observations of drag and velocity profile have also been used to determine the variation of mixing length with height. In unstable conditions the mixing length increases more rapidly than the height, while in stable conditions the increase is likely to be less rapid than the height.

INTRODUCTION

In any theory of turbulent fluid motion the relation of the drag exerted by the bounding surface to the mechanics of the motion is fundamental. In pipe-flow and in the wind-tunnel drag measurements are fairly straightforward, and the data now

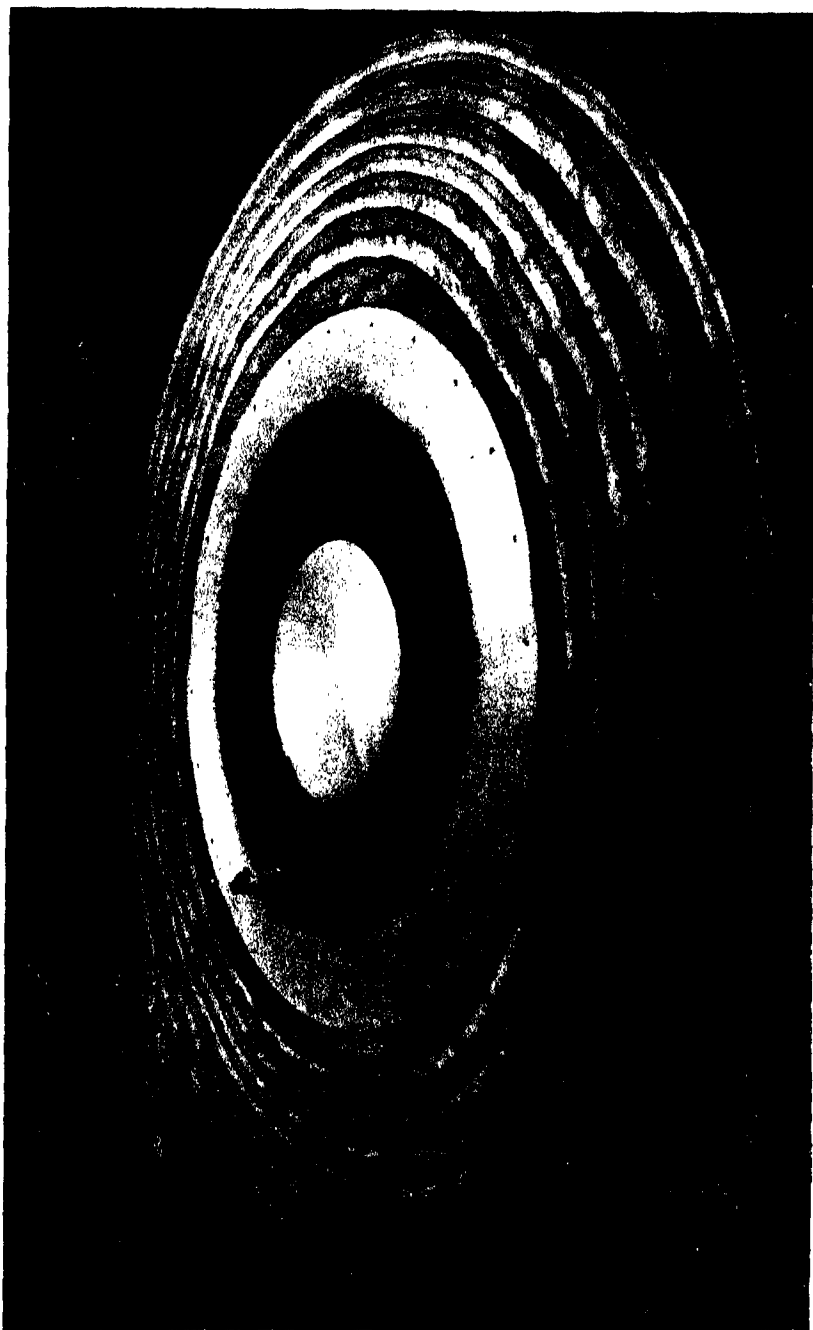


FIGURE 1. Surface drag-plate in oil bath and fair-d surround.

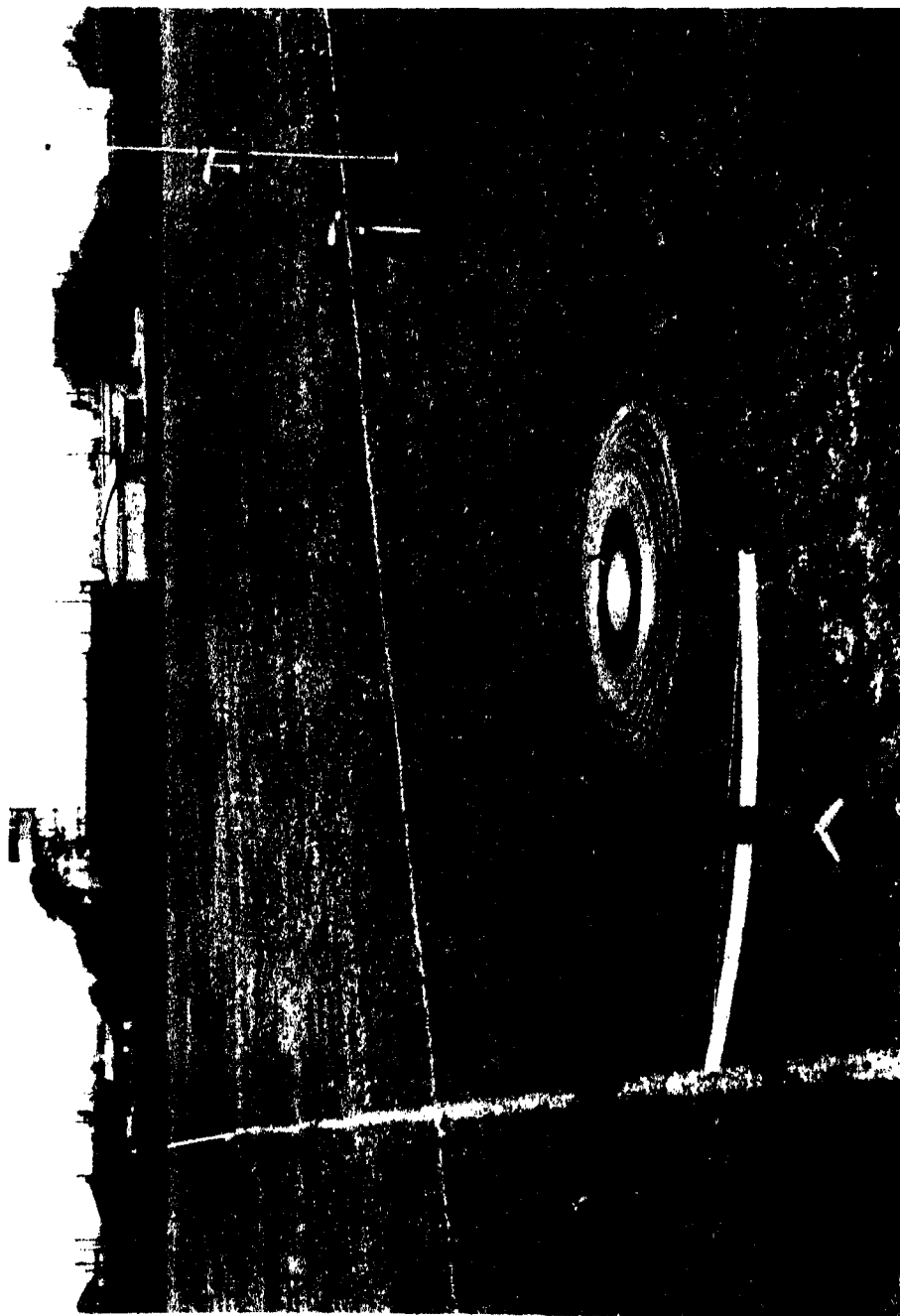


FIGURE 2. Disposition of surface drag-plate and associated equipment in the field including anemometers
(right) for measurement of velocity profile.

available have been widely used to check theoretical treatment (see, for example, Nikuradse 1932, 1933; Prandtl 1935; Sutton 1934).

The drag which the earth's surface exerts on the wind has not yet, however, been directly measured,* though it is of importance, in itself, as a source of dissipation of the energy of the wind, and as providing a measure against which to check theoretical treatment for a scale of flow much greater than is encountered in hydraulics and aeronautics. The present investigation was therefore undertaken to make good the previous absence of data on the drag of the earth's surface and to provide for the first time a measure of von Karman's constant for the lower atmosphere, since this constant occupies a central position in modern theories of turbulence.

The lower atmosphere is of special interest in problems of turbulent flow, not merely on account of scale but because of the marked vertical stability or instability which is frequently present. The effects of stability on mixing-length theories of turbulence will therefore be considered in the light of the present observations and of those of earlier workers.

RELATIONS BETWEEN DRAG, MIXING LENGTH AND VON KARMAN'S CONSTANT

The mixing-length theory of turbulence has been developed along lines of somewhat intuitive reasoning, and no attempt will be made here to justify the processes or assumptions.

The horizontal shear stress τ due to turbulence in horizontal two-dimensional flow is given by the Reynolds equation

$$\tau = -\rho \overline{u'w'}, \quad (1)$$

where ρ is the fluid density, u' , w' are the eddy velocities in the downwind and vertical directions respectively, and the bar denotes a time-mean value of the product.

The concept of a mixing length l which is the mean length traversed by turbulent elements before mixing completely with their surroundings is now introduced. Prandtl sets $u' = l' \partial u / \partial z$, where u is the mean horizontal velocity at height z and l' the mixing path for the particular turbulent element, and he further takes $w' \propto u'$. Then by eliding the latter constant of proportionality with l' (thus adding to the lack of exact physical definition of l) (1) becomes

$$\tau = \rho l^2 (\partial u / \partial z)^2. \quad (2)$$

In the steady state τ is necessarily independent of height (when $\partial u / \partial t = 0$, $\partial \tau / \partial z = 0$ also, from Fick's equation), but in fact it may always be taken as constant in the first few metres of the atmosphere (Calder 1939).

* Except in a rough-and-ready way by L. F. Richardson (1922), who observed the deflexion of corn stalks in a wind and measured the force necessary to produce the observed deflexion.

Von Karman (1930), making the assumption of an identical pattern of turbulent flow, apart from scale, throughout the field relates l with the field of mean velocity by

$$l = k \frac{|\partial u / \partial z|}{|\partial^2 u / \partial z^2|}, \quad (3)$$

where k is von Karman's constant. This relation introduced in (2) gives

$$k = \sqrt{\left(\frac{\tau_0}{\rho}\right) \frac{|\partial^2 u / \partial z^2|}{(\partial u / \partial z)^2}},$$

where τ_0 replaces τ , implying the shear at the surface, $z = 0$.

Thus, measurements of the surface drag and of the velocity profile will determine k .

The last equation will be applied in its integrated form for a fluid flowing over a hydrodynamically rough surface (completely developed turbulence),* viz.

$$u = \frac{1}{k} \sqrt{\left(\frac{\tau_0}{\rho}\right)} \log \frac{z + z_0}{z_0}, \quad (4)$$

where z_0 is a constant of integration and $u = 0$ at $z = 0$. If $u = u_1$ at a reference height z_1 , (4) gives

$$u_1 - u = \frac{1}{k} \sqrt{\left(\frac{\tau_0}{\rho}\right)} \log \frac{z_1 + z_0}{z + z_0}, \quad (5)$$

but, since z_0 for the surface over which observations have here been taken was of order 10^{-3} to 10^{-4} cm. as deduced from observations of wind at two or more levels, whereas the wind velocity was measured at heights from about 20 to 200 cm., the formula for k may, with sufficient accuracy, be written

$$k = \sqrt{\left(\frac{\tau_0}{\rho}\right) \frac{\log z_1/z}{u_1 - u}}. \quad (6)$$

This is the relation which has been used in the present investigation for the determination of k . The same formula is valid if the underlying surface is aerodynamically smooth and the velocity profile given by

$$u = \frac{1}{k} \sqrt{\left(\frac{\tau_0}{\rho}\right)} \left\{ \log \left(\frac{z}{\nu \sqrt{\left(\frac{\tau_0}{\rho}\right)}} \right) + \text{const.} \right\}$$

instead of (4), where ν = kinematic viscosity of fluid (see Prandtl 1935, p. 133).

Nikuradse (1932) first determined von Karman's constant from measurements of the resistance to the flow of water through a cylindrical pipe in which the velocity distribution was simultaneously measured. He obtained a value $k = 0.40$, in-

* A surface is hydrodynamically rough and the turbulence fully developed if the fluid boundary layer is turbulent right up to the boundary surface. The criterion for such flow is that $\epsilon/\nu\sqrt{(\tau_0/\rho)} > 5$ (Nikuradse 1933), where ϵ is the height of the roughness elements composing the surface and ν the kinematic viscosity of the fluid. In neutral equilibrium the velocity profile may also be used as a criterion; if the observations conform with equation (4) the turbulence is fully developed.

dependent of Reynolds's number, and one may expect it to be a universal constant having the same value for all fluids and all forms of isothermal flow. It was, in fact, for isothermal flow that the mixing-length theory was formulated and the effect of buoyancy was therefore excluded. There is, however, nothing in the concept of mixing length which precludes an extension to more general conditions, though considerations will later be advanced (see p. 217) regarding the likely variation of k under conditions of thermal gradient, i.e. stable or unstable stratification.

It is noted from (4) that the drag in fully developed turbulent flow is proportional to u^2 , independent of Reynolds's number. If a drag coefficient c_D is introduced for the earth's surface analogous to that used in aerodynamics, then

$$\tau_0 = \frac{1}{2} c_D \rho u_1^2,$$

where now c_D is necessarily related to the reference level at which the velocity u_1 is measured. Then, using this equation with (4) it is found that c_D is otherwise a function only of k and of the roughness of the boundary, i.e. of z_0 , according to the relation

$$c_D = \frac{2k^2}{\left(\log \frac{z_1 + z_0}{z_0}\right)^2}. \quad (7)$$

It may be readily inferred from (2) and (4) above that

$$l = k(z + z_0), \quad (8)$$

a relation that will be tested by the observations to be presented and compared with Nikuradse's results.

APPARATUS AND METHODS OF OBSERVATION

Measurement of drag. It is not possible to measure the drag of the earth's surface from measurements of the horizontal-pressure field without making very doubtful assumptions, and one is therefore forced to more direct but more troublesome methods. The ideal apparatus for such a direct method would consist of an area of the natural ground surface freed from its surroundings in such a way that the tangential force upon it could be measured without movement or distortion of the area or of its surroundings and without disturbing the structure of the wind. An approach to this ideal has been made by floating a flat plate in liquid contained in a very shallow bath with a wide, faired surround (figure 1, plate 1) set down on an extensive surface of smoothness approximating to that of the floating plate. The latter was under the control of a vertical torsion suspension carried in a frame fixed to the bath and connected by an arm dipping below the liquid surface to the perimeter of the plate. The apparatus was orientated so that the arm was approximately at right angles to the wind direction, and the deflexion of the plate by the wind was observed by means of a distant telescope and scale and a mirror carried on the torsion suspension. Figure 2, plate 2, shows the complete apparatus as used in the field, though not

on grass as in this figure. The zero position of the drag-plate was observed by covering the bath with a draught-proof screen in which a slot allowed the mirror to be visible. A clamp, not shown in figure 1, was used to hold the plate and to relieve the suspension of strain during the transport and setting up of the apparatus in the field.

Details of the method of suspension and of drag-plate construction are shown in figure 3. The face of the plate, 22.5 cm. diameter, was of aluminium with a smoothly turned rim to which cellophane was attached to form the under-side. The bath was 45.4 cm. diameter, and the distance from the plate centre to the suspension 20.7 cm. The plywood surround was 107 cm. diameter and the total depth of the apparatus only 3.5 cm., so that the air flow was practically undistorted. Controls were provided on the suspension head and frame for (i) elevation of the arm of the drag-plate, (ii) tension in the suspension wire, (iii) zero. The suspension was of 36 s.w.g. phosphor-bronze wire which was found to be entirely suitable, free from hysteresis and creep, and preferable to other materials such as quartz.

Various liquids were tried for use in the bath, the choice ultimately falling on 'instrument oil' which had the appropriate viscosity (about 1 poise at 15°C) to provide critical damping and a low surface tension ensuring a free motion and stable zero for the plate. The bath was filled to the brim with oil to avoid discontinuities in the air flow at this point, and the transition at the edge of the plate was also smooth. A strong wind set the oil into rippling motion and made the drag-plate unsteady, so that observations were confined to winds of less than about 6 m./sec. (15 miles/hr.).

The torsion constant c of the apparatus was measured by placing small weights up to 0.45 g. in a paper pan connected to the drag-plate by fine cotton passing over a light pulley. The resulting calibration curves relating the angular deflexion θ with the applied moment were linear, with negligible scatter of points and with good agreement between θ increasing and θ decreasing. The mean value of c for three calibrations was 216 dyne-cm. per degree with 1 % spread.

The value of τ_0 required for equation (6) is then connected with the deflexion of the drag-plate in the field by the relation

$$\tau_0 AL \sin \alpha = c\theta, \quad (9)^*$$

* *Footnote added 13 May 1946.* This equation neglects (i) the restoring force arising from the banking up of the oil on the leeward side of the bath, and (ii) the viscous force in the direction of τ_0 due to the motion of oil across the bath at and near the upper surface. The inclination of the upper surface of the oil in a wind of 5 m./sec. at 2 m. height was too small to be observed, while theory well supported by observations in shallow seas (Sverdrup 1945), gives, in conjunction with the results of table 1 below, a value of 10^{-4} rad. The *observed* surface speed of the oil in the above wind was 0.1 cm./sec. in the direction of the wind. The force per unit area of plate arising from the inclined surface would then be about 0.05 dyne/cm.², whereas that due to the moving oil, mainly exerted on the lower face of the drag-plate and equal to $\mu du/dz$, the symbols relating to the oil, would also be about 0.05 dyne/cm.², but in the opposite direction. The observed drag in a wind of 5 m./sec. as given by equation (9) was 0.31 dyne/cm.² (see table 1), and equation (9) is therefore considered to be sufficiently exact.

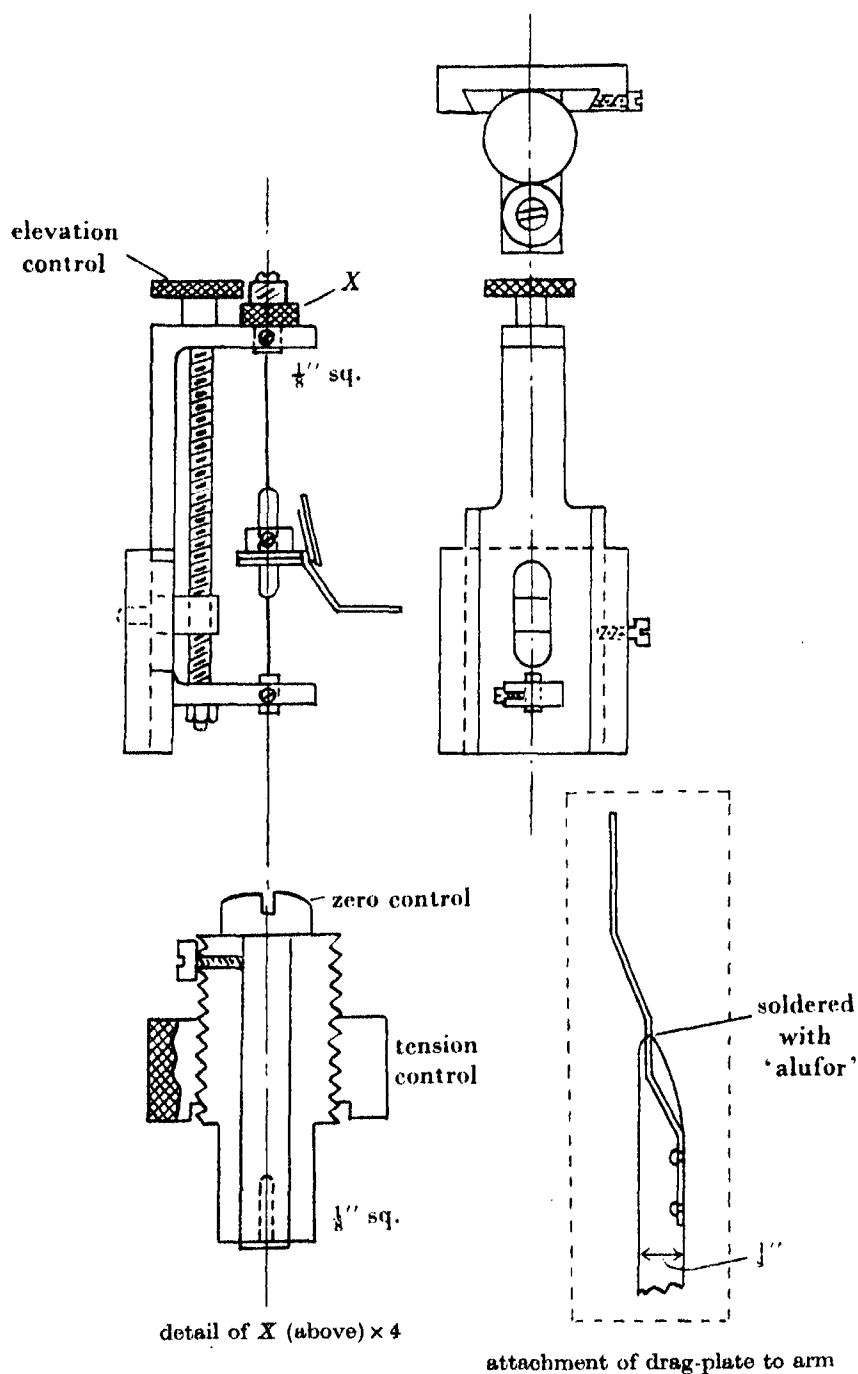


FIGURE 3. Suspension and frame for drag-plate.

where A is the area of the plate, L the distance from the suspension to the centre of the plate, and α is the angle between the wind direction and the line of symmetry of the plate and suspension.

Velocity profile. The mean wind velocity was determined over periods of 10 min. duration at heights of approximately 20, 50, 100 and 200 cm. above the surface on which the drag-plate was mounted by means of Sheppard (1940) portable three-cup anemometers. The mounting of three of these instruments is to be seen at the right of figure 2, plate 2.

Field technique. It was essential that the drag-plate should be used on an area of ground whose surface was closely similar to that of the plate itself. This was available in the form of a horizontal circular area of concrete 160 m. in diameter located on a ridge on Salisbury Plain. The surface of the concrete was smooth in general appearance with small-scale roughness of the order of 1 mm. in dimensions. At some points of its circumference the concrete was level with, and at others slightly below, the contour of the downland immediately surrounding the area, while at a little greater distance from the perimeter the land fell away rather steeply in most directions. These features made it necessary to provide the maximum possible distance between the upwind edge of the concrete and the site of the drag-plate and wind-profile equipment. The distance was in fact well over 100 m., which was sufficient to ensure a boundary layer to a height of 2 m. (the maximum height of the anemometers) characteristic of the plane concrete surface except in the more stable atmospheric stratifications. The lack of complete adjustment of the wind at the downwind end of the concrete in stable conditions made it, in fact, impossible to obtain reliable observations under these conditions, though some indication of the changes of drag, etc., accompanying increasing stability was obtained.

It had been hoped to modify the drag-plate so that the apparatus could be used on downland turf, thus avoiding the above difficulties of exposure. But the investigation had to be discontinued before these changes could be made, and the results are accordingly more limited in range than is desirable in a meteorological investigation.

The anemometers, drag-plate and associated equipment having been set up on the concrete along a line normal to the wind direction to avoid mutual interference, observations were taken over 10 min. periods. Since the natural wind is turbulent the deflexion of the drag-plate changed continuously, and it was accordingly observed every 6 sec. through each period and a mean value deduced from the hundred observations. Using bars to denote mean values, $\bar{\tau}_0$ was calculated from the following equation, a modification of (9),

$$\bar{\tau}_0 A L \sin \bar{\alpha} = c \bar{\theta}, \quad (10)$$

in which $\bar{\alpha}$ was determined by taking a record of the variations in wind direction at the site throughout any period of observation by means of a light vane (not shown in figure 2, plate 2), a compass bearing being later taken on the centre of the trace and on the line of symmetry of the plate at a given value of θ .

It was found that the probable error of the arithmetic mean of θ , as defined for a normal error distribution, was just over 1 %. In fact, the distribution of θ was slightly skew so that the probable error of the mean can have no strict significance, but the divergence from the normal error law was slight, and 1 % can still be taken as a rough guide of the error involved in the measurement of $\bar{\theta}$ and of $\bar{\alpha}$ which differs from $\bar{\theta}$ by a constant quantity.

But equation (10) is not exact except when $\bar{\alpha} = 90^\circ$ and $\sin \alpha$ symmetrical, therefore, about the mean. In fact, it was generally possible to orient the drag-plate initially so that $\bar{\alpha}$ should not differ greatly from 90° , and in the three runs reported below this was the case. The smallest instantaneous value of α was 53° , for which $\sin \alpha (= 0.80)$ is inside the range of comparatively slow variation. The error in k (which is proportional to $\sqrt{\tau_0}$) arising from the approximation of (10) and from the method of measuring $\bar{\theta}$ is considered to be not more than 2 or 3 %.

The continuous record of wind direction referred to above served a subsidiary purpose; it enabled any change in mean wind direction to be noticed. When such a change occurred during a run the results were discarded for equation (10) was not then valid.

RESULTS AND GENERAL DISCUSSION

Completely satisfactory observations were obtained on three runs carried out under conditions of slight instability in winds of about 4 m./sec. For each run k has been evaluated from equation (6) over three height intervals corresponding to the four levels of observation of wind speed, and the results are presented in table 1. This table also includes the drag coefficient c_D for a reference level for velocity of 2 m. and temperature difference ($T_{710} - T_{120}$) observed between 710 and 120 cm. on a recording installation (Johnson 1929) on downland about a mile away from the site of the present experiments. On account of the difference in thermal

TABLE 1. OBSERVATIONS OF DRAG, VELOCITY PROFILE AND
VON KARMAN'S CONSTANT ON 26 APRIL 1939

(Subscripts refer to heights in cm.)										
run num- ber	time (a.m.t.) weather	$u_{18.4}$ (cm./ sec.)	$u_{48.8}$ (cm./ sec.)	u_{102} (cm./ sec.)	u_{202} (cm./ sec.)	τ_0 (dynes/ cm. ²)	ρ (g./cm. ³)	c_D for u $= u_{200}$	mean k	$T_{710} - T_{120}$ (°C)
1	13.35-45 C., Sc., 8/10 Cs., 8/10	405	449	480	498	0.310	1.23×10^{-3}	2.0×10^{-3}	0.45	-0.8
2	14.40-50 C., Sc., 6/10 Ci., Cs., 8/10	328	358	377	398	0.206	1.23×10^{-3}	2.1×10^{-3}	0.44	-0.8
3	14.53 15.03 C., Sc., 8/10 Ci., 9/10	310	333	356	366	0.193	1.23×10^{-3}	2.3×10^{-3}	0.49	-1.0
	mean	348	380	404	421	—	—	2.15×10^{-3}	0.46	-0.9

characteristics of grass land and concrete and of the distance apart of the two sites these values can only be taken as a rough measure of the thermal instability of the air over the concrete.

The drag coefficient c_D is observed to be approximately constant. Such variation as is found may be partly real and due to the effect of varying stability or to a non-square law of drag for a surface which may not have been aerodynamically rough (in this case $c_D \propto u_1^{-0.2}$ approximately), but some of the variation must be attributed to experimental error.

The individual and mean values obtained for k are in encouraging agreement with Nikuradse's (1932) value of 0.40 for pipe flow when it is borne in mind that the moderate instability shown in table 1 might be expected to lead to somewhat higher values than in the isothermal flow of the laboratory. (Instability must produce greater mixing lengths than isothermal conditions, and equation (8) then implies an augmented value of k .) The measure of agreement found is not, of course, surprising in view of the anticipated generality of von Karman's constant, but the result is reassuring bearing in mind the difference in fluids and the tremendous difference in scale of the two motions.

The immediate significance of the result is that the drag of any natural surface over which the air is in near-neutral equilibrium and in which turbulence is fully developed may be deduced immediately by observing the velocity profile and inserting $k = 0.40$ in equation (7). It is thus assumed that the small difference between Nikuradse's and the value of k given here is due to the moderate instability of the air in the present investigation and perhaps in part to experimental error.

The drag coefficient given in table 1 above is not, of course, representative of the values pertaining to most naturally occurring surfaces, except perhaps a fairly smooth sea, since an unusually smooth surface was used. Having, however, established the

TABLE 2. VALUES OF c_D AS FUNCTION OF u_{2m}/u_{1m} OR OF z_0
IN CONDITIONS OF NEUTRAL EQUILIBRIUM

(1)	(2)	(3)	(4)
u_{2m}/u_{1m}	z_0 (cm.)	type of surface giving values shown in (1) and (2)*	drag coefficient c_D (using velocity at 2 m. as reference velocity)
1.06	0.0009	one with rather smooth appearance, e.g. ice or smooth snow	2.11×10^{-3}
1.10	0.10	lawn grass up to 1 cm. long	5.54×10^{-3}
1.14	0.72	downland; thin grass up to about 10 cm.	1.01×10^{-2}
1.18	2.3	thick grass up to about 10 cm.	1.60×10^{-2}
1.22	5.0	thin grass up to about 50 cm.	2.32×10^{-2}
1.26	8.9	thick grass up to about 50 cm.; thin wheat, etc., up to about 100 cm.	3.21×10^{-2}
1.30	13.9	fully grown root crops	4.28×10^{-2}

* The notes in column (3) are based on Paeschke (1938) and unpublished observations by the author.

validity of equation (7) with $k = 0.4$ for the atmosphere, table 2 gives values of c_D for various values of the velocity ratio u_{zm}/u_{1m} , and corresponding values of z_0 which are liable to be found over natural surfaces of varying roughness. In measuring the velocity profile and hence z_0 for surfaces whose roughness elements are an appreciable fraction of the height at which the velocity is measured, a zero-point displacement equal to a major fraction of the height of the roughness elements should be made in applying equation (4) (see Paeschke 1938).

It remains to point out that a surface may be aerodynamically rough for moderate and high wind velocities, so allowing the application of equation (7), but it may become aerodynamically smooth at low wind velocities (see footnote, p. 210) and another law then applies. But the only natural surfaces likely to be concerned are smooth ice and water. In the latter case the wind itself determines the geometrical roughness to a large degree and still further complicates the problem (see Rossby & Montgomery 1936).

EFFECT OF ATMOSPHERIC STABILITY ON k

It has already been mentioned that it had been hoped to investigate quantitatively the variation of k with atmospheric stability, but that on account of the limited extent of the concrete surface the wind profiles at the experimental site were unrepresentative of the boundary layer over the surface under stable stratification. Nevertheless, some indication of the behaviour of k under such conditions was obtained.

Equation (6) may be written

$$k = \frac{\sqrt{(\tau_0/\rho)}}{u_1} \times \frac{\log z/z_1}{1 - u/u_1},$$

where $z < z_1$ ($u < u_1$). The ratio u/u_1 decreases with increasing stability (Best 1935, and others), so that the second factor on the right of this equation decreases also.

If then the quantity $\frac{\sqrt{(\tau_0/\rho)}}{u_1}$ is observed to decrease as the vertical temperature gradient changes from lapse to inversion, k must decrease still more rapidly, and a lower limit to the variation in k is determinable from the behaviour of this term alone.

A number of runs were made under such conditions and in all cases $\frac{\sqrt{(\tau_0/\rho)}}{u_1}$ was observed to decrease whichever anemometer height z_1 was taken as standard, though the irregularities of the profile prohibited a worthwhile computation of k itself. During these observations the temperature gradient between 23 and 135 cm. was measured over the concrete itself, and the results of one series of observations during which the temperature difference ($T_{135} - T_{23}$) increased steadily by 0.5°C over a 3 hr. period in the evening may be quoted. The percentage decrease in $\frac{\sqrt{(\tau_0/\rho)}}{u_1}$ for $z_1 = 200$, 100, 60, 40, 25 and 20 cm. (the number of levels of anemometer observation had been increased in these later runs) was 18, 21, 14, 15, 28 and 22 % respectively, i.e. mean decrease of 20 %.

Thus there is evidence for a material decrease in k as atmospheric stability increases. Such a statement only has meaning, however, if the concepts giving rise to k , e.g. equation (8), are valid in all thermal stratifications, and in particular the logarithmic velocity profile must be generally valid. This is not, in fact, precisely true, but departures from the logarithmic law are only considerable for appreciable departures from neutral equilibrium (see p. 221), and one waits for a physically satisfying theory which will include these conditions.

There is a second, again somewhat approximate, method of investigating the variation of k with stability. Sutton's theory of turbulence, making use of the mixing-length concept and of von Karman's constant, leads to a power law of wind structure in the lower layers of the atmosphere

$$u/u_1 = (z/z_1)^p \quad (0 \leq p \leq 1).$$

This relation is less accurate than the logarithmic law for neutral equilibrium and low values of z , but it is probably a better representation in inversions for all z and is a good approximation in all stabilities for z greater than about 1 m. Introducing the derivatives of the velocity given by this power law into equation (3) and setting

$$l = \frac{|w'|}{\partial u / \partial z},$$

after Prandtl, in the same equation, then*

$$\frac{|w'|}{u} = k \left| \frac{p}{p-1} \right|. \quad (11)$$

Now Best has simultaneously measured wind profiles to 2 m. and the vertical gustiness by means of a modified Taylor bi-directional vane, in a complete range of measured temperature gradients. The observations with the bi-directional vane give in fact a quantity proportional to $|w'|/u$, but the factor of proportionality, the ratio of $|w'|$ to the effective maximum of $|w'|$, is rather uncertain. If, however, $k = 0.4$ is set for neutral equilibrium (zero-gradient of temperature over the height range covered by Best determines this condition with quite sufficient accuracy) this enables equation (11) to be used to determine the factor of proportionality and hence to determine k in other temperature gradients. Best's data have been analyzed along these lines, taking p from the winds at 2 and 1 m., and the results are shown in figure 4. This curve confirms the inference made above as to the considerable variation of k with temperature gradient. (The analysis gave the factor of proportionality referred to above as 0.28, a very reasonable value.)

The results of this section indicate that the drag of the earth may be obtained for any thermal stratification provided that a value of k appropriate to the stratification is used in equation (7), and provided that the velocity profile enables a sound value of z_0 to be obtained. Since c_D varies as k^2 the effect of stratification as shown in

* This relation was first noticed by Mr K. L. Calder and the author in a joint investigation which has not been published.

figure 4 is considerable. That is not to say, however, that c_D , for a fixed height of reference, varies considerably with the stratification; the variation of c_D is, indeed, surprisingly small as shown by table 3, which has been built up on Best's simultaneous measurements of the temperature gradient and velocity profile over a sports field (grass length about 1 cm.) and from figure 4. (Best's profiles show good agreement with the logarithmic law in each of the temperature gradients listed in table 3.) But had $k = 0.4$ been assumed throughout, the corresponding values of c_D would have been 2.3 , 7.4 and 10.3×10^{-3} for lapse, zero gradient and inversion respectively, i.e. a fivefold increase instead of the approximate constancy shown in table 3. For a given surface pressure gradient the wind at 2 m. falls off very considerably in passing from lapse to inversion in the lower atmosphere, so that the drag itself would in general show an appreciable decrease with the above increase in stability.

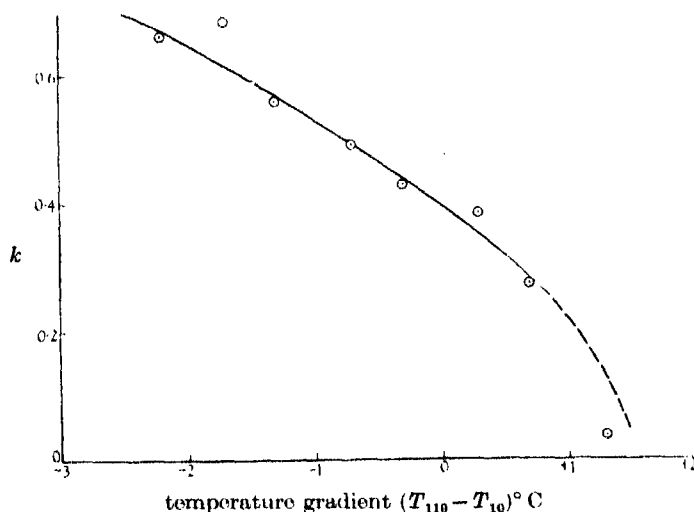


FIGURE 4. Variation of k with atmospheric stability.

TABLE 3. VARIATION OF DRAG COEFFICIENT FOR A SHORT-GRASS SURFACE WITH ATMOSPHERIC STABILITY

temperature gradient (°C/m.)	z_0 (cm.)	k from figure 4	c_D for reference height of 2 m.
-1.67	0.0057	0.61	5.3×10^{-3}
0.0	0.28	0.40	7.4×10^{-3}
+0.56	0.80	0.22	3.1×10^{-3}

MIXING LENGTH AND ITS VARIATION WITH HEIGHT

The mixing length is given by equation (2) with the assumption of shearing stress independent of height as

$$l = \frac{\sqrt{(\tau_0/\rho)}}{\partial u/\partial z}. \quad (12)$$

l may therefore be deduced as a function of height from the results given in table 1 without assuming the form of the velocity profile to be logarithmic. If, in fact, the logarithmic law were accurate points on the l, z diagram would lie on a straight line given by (8) and the slope would be the measure of k .

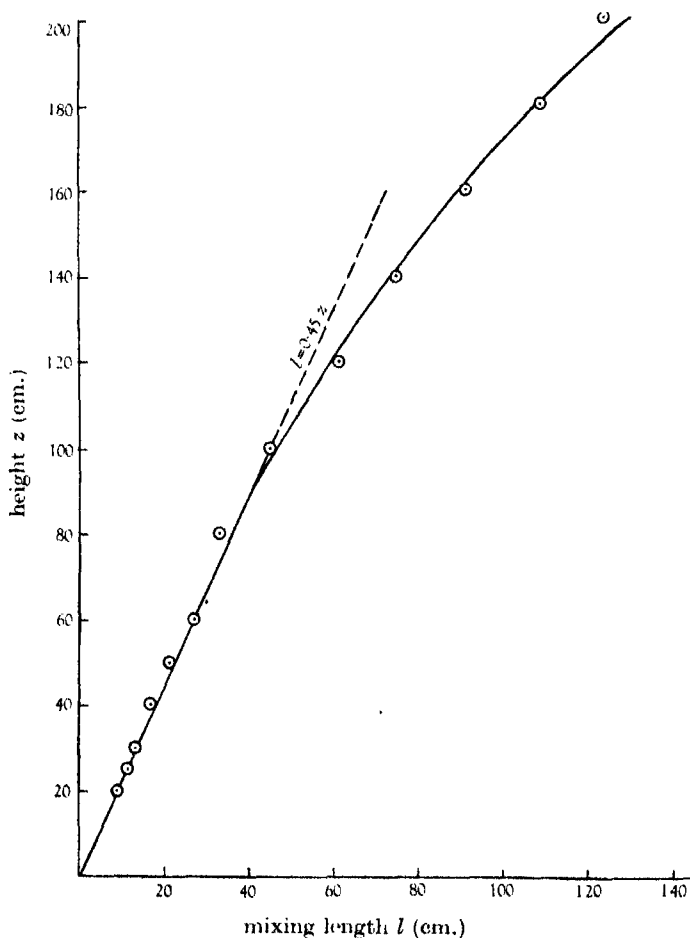


FIGURE 5. The variation of mixing length with height.

The observations of $\sqrt{\tau_0}$ and of the velocity distribution given in table 1 have been meaned from the three runs there reported, and, using a tangent measure on the mean profile, values of l have been deduced at the levels shown by the points of figure 5. It is observed that the points fit the line $l = 0.45z$ up to a height of about 1 m., with as little scatter as can be expected from a technique involving the use of tangents. Above that level l increases more rapidly than the height, implying some departure from the logarithmic law for wind speed at those levels. Further, the value 0.45 of the slope of the linear portion is in good agreement with the value of k deduced from (6).

Power laws have been much used in aerodynamic investigations (often as approximation formulae), and a $\log z, \log l$ plot of all points in figure 4 shows that the l, z variation is quite well represented over the whole range by

$$l = 0.25z^{1.15}.$$

There is now good evidence (Firesah & Sheppard, not yet published) that the logarithmic velocity profile of equation (4) is in excellent accord with observation in the atmosphere over many kinds of surfaces, provided the atmosphere is in neutral or near-neutral equilibrium (potential temperature constant or varying only very slowly with height), but that in appreciable lapses of temperature the $u, \log z$ curve is convex to the u -axis (an example is found in the present observations) and in inversions of temperature it is concave to the u -axis. It follows then from equation (12) *that in lapse conditions the mixing length increases more rapidly than the height, while in inversions the increase is less rapid than the height.*

It is not possible, at present, to provide a physical explanation of the power-law variation of mixing length with height; such a law is, indeed, likely to be inexact, suitable only for interpolation purposes and restricted as to height. But it presents a challenge in respect of the variation of the wind profile with stability which it is hoped to follow up in another paper. Meanwhile, one important consequence of the variation of profile with stability which has been referred to above may be noted. The eddy viscosity K , which is of profound importance in exchange processes in the lower atmosphere, is given by

$$\tau = K\rho\partial u/\partial z,$$

which, with $\partial\tau/\partial z = 0$ gives

$$K \propto (\partial u/\partial z)^{-1},$$

and the observed variation in the wind profile as between lapse and inversion of temperature implies that K increases more rapidly than the height in the former condition and less rapidly than the height in the latter. Only in neutral equilibrium is $K \propto z$.

The results of this section imply that the k concept must be used with care in the lower atmosphere, and that inaccuracies may accrue in computations involving either major instability or stability even when k is adjusted, e.g. from figure 4, for the particular stratification.

The writer is very pleased to acknowledge the helpful advice of Mr L. F. G. Simmons of the National Physical Laboratory in regard to the technique of drag measurement and of Mr H. Scamell of the Ministry of Supply, in the details of design of the drag apparatus. He is also much indebted to Mr C. H. B. Priestley, now of the Meteorological Office, who assisted, with others, in the field observations and with whom the author has had helpful discussion of the results.

Acknowledgement is also gratefully made to the Ministry of Supply for permission to publish the experimental work described in this paper.

REFERENCES

- Best, A. C. 1935 *Geophys. Mem.* **65** (7, 8).
 Calder, K. L. 1939 *Quart. J.R. Met. Soc.* **65**, 537.
 Johnson, N. K. 1929 *Geophys. Mem.* **46** (5, 6).
 Karman, Th. von 1930 *Nachr. Ges. Wiss. Göttingen*, p. 58.
 Nikuradse, J. 1932 *Forschungsh. Ver. deutsch. Ing.* no. 356.
 Nikuradse, J. 1933 *Forschungsh. Ver. deutsch. Ing.* no. 361.
 Paeschke, W. 1938 *Beitr. Phys. frei. Atmos.* **24**, 163.
 Prandtl, L. 1935 *Aerodynamic Theory, III. Div. G.* (ed. Durand). Berlin: Julius Springer.
 Richardson, L. F. 1922 *Weather Prediction by Numerical Process*, p. 85. Cambridge University Press.
 Rossby, C.-G. & Montgomery, R. B. 1936 *Pap. Phys. Oceanogr. Met., Mass. Inst. Tech.* **4**, 3.
 Sheppard, P. A. 1940 *J. Sci. Instrum.* **17**, 218.
 Sutton, O. G. 1934 *Proc. Roy. Soc. A*, **146**, 701.
 Sverdrup, H. U. 1945 *Oceanography for Meteorologists*, p. 121. London: Allen & Unwin.

The structure of polyisoprenes. VI

An investigation of the molecular structure of dibenzyl by X-ray analysis

BY G. A. JEFFREY

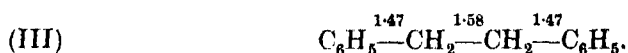
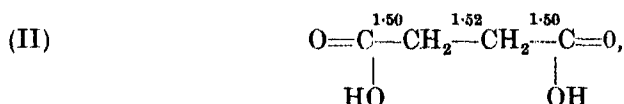
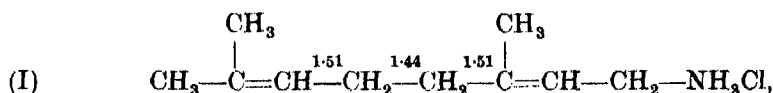
(Communicated by W. N. Haworth, F.R.S.—Received 27 July 1945)

The molecular structure of dibenzyl has been determined to an accuracy of about 0.01 Å by using extensive X-ray data in three-dimensional Fourier syntheses to refine the approximate atomic parameters originally derived by Robertson. The three formally single-carbon bonds joining the benzene rings have lengths of 1.50, 1.48 and 1.50 Å and make angles of 115° with each other. The departure from the standard length of 1.54 Å is discussed in relation to the analogous systems of polyisoprenes and 1,5 dienes where related effects have been observed. The dimensions of the benzene rings also reflect the unusual character of the acyclic carbon bonds, two of the aromatic bond lengths being 1.39 Å and the other four 1.37 Å. The molecule, although possessing a centre of symmetry, is not absolutely symmetrical, since the central $\text{CH}_2\text{--CH}_2$ bond is inclined at 70.5° to the plane of the benzene rings. This is ascribed to the influence of intermolecular forces.

INTRODUCTION

The first detailed structural examination of a crystalline compound of the polyisoprene type represented by geranylamine hydrochloride (I) revealed an anomalous interatomic distance for the bond linking the two isoprene units. This clearly called for further investigation both to confirm an isolated observation and, if it was found that a general principle was involved, to establish the molecular conditions under which it functioned. An attempt at correlation with the ultra-violet absorption

spectra by Bateman & Koch (1944) gave negative results, presumably because the structural feature reflected no abnormality in that absorption region. Neither did the chemistry of the simpler members of the same class of compound offer any decisive evidence, although for certain carbathoxyl-substituted hexadienes the assumption of special properties for the 1.5 system of double- and single-carbon bonds was shown by Bateman & Jeffrey (1945) to provide an explanation of some exceptional chemical properties.



From structural analyses, either by X-rays or electron diffraction, only two examples could be found where related compounds had been examined in detail. These were the crystal structures of succinic acid (II) by Verweel & MacGillavry (1939) and of dibenzyl (III) by Robertson (1935). In both cases it was doubtful whether the results were sufficiently precise to have any bearing on the problem. Of the two, the dibenzyl was chosen for further inquiry using the more powerful modern X-ray methods because the revision of these bond lengths was the more necessary, if only to confirm or otherwise the existence of a covalent carbon to carbon bond longer than that of diamond which has also been reported for hexamethyl ethane by Bauer & Beach (1942*a, b*).

SUMMARY OF ROBERTSON'S STRUCTURE ANALYSIS

A unit cell of dimensions $a = 12.77 \text{ \AA}$, $b = 6.12 \text{ \AA}$, $c = 7.70 \text{ \AA}$, $\beta = 116^\circ$ was found to contain two centrosymmetric molecules of $\text{C}_6\text{H}_5 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{C}_6\text{H}_5$ related by the space-group symmetry $2_1/a$. The phase angles of the structure factors are therefore multiples of $\frac{1}{2}\pi$, and their derivation together with the experimental data leads to the determination of the twenty-one unknown atomic parameters excluding the hydrogen atoms.

About 150 absolute intensities, mainly of planes in the principal zones, were measured experimentally using a two-crystal moving-film spectrometer and a Robinson integrating photometer. From these the approximate orientation of the benzene rings with respect to each other and the crystal axes were deduced by trial and error methods, and the co-ordinates were refined by Fourier projections on the principal planes. The (100) projection was of little value owing to overlapping of adjacent molecules in the cell, but a fair degree of resolution was obtained in the other directions. The benzene rings were shown to be approximately planar and

TABLE 1. COMPARISON OF STRUCTURE AMPLITUDES

R , experimental $|F|$ values measured by Robertson. J , experimental $|F|$ values obtained in present work. C , theoretical F values calculated from final co-ordinates.

hkl	R	J	C	hkl	R	J	C	hkl	R	J	C
200	38	47	-47	405	3	4	4	310	15	16	15
400	7	7	13	404	< 3	< 2	- 3	320	18	17	-18
600	16	16	16	403	< 3	2	2	330	< 3	1	0
800	14	13	13	402	34	33	30	340	< 6	< 2	- 1
1000	11	10	- 9	401	25	24	18	350	< 6	2	3
				40 $\bar{1}$	12	13	12				
020	10	10	4	402	3	4	- 9	410	< 4	3	1
040	9	9	- 9	403	< 2	2	3	420	25	24	25
060	< 3	2	- 2	404	< 2	< 2	2	430	10	10	- 9
				405	< 2	< 2	0	440	< 3	< 2	- 1
001	20	19	19	406	11	11	11	450	< 6	3	- 4
002	20	24	-21	407	< 3	< 2	0				
003	7	7	10					510	16	17	13
004	12	12	- 9	604	6	4	7	520	5	6	4
005	3	3	4	603	6	6	5	530	< 5	4	3
006	< 3	< 2	- 1	602	2	2	0	540	< 6	< 2	1
				601	2	2	3	550	< 6	3	- 3
011	59	59	-62	60 $\bar{1}$	11	10	10				
012	18	17	13	602	18	19	-18	610	< 5	3	- 5
013	3	4	- 4	603	4	4	- 3	620	9	10	- 8
014	4	4	4	604	13	14	11	630	< 6	< 2	1
015	5	5	4	605	< 3	< 2	0	640	< 6	6	- 6
016	< 3	< 2	- 1	606	9	9	-10				
				607	< 3	2	- 2	710	6	6	- 8
021	19	16	18					720	< 5	2	- 3
022	6	6	- 2	803	< 3	< 2	0	730	< 6	2	- 3
023	3	3	- 2	802	< 3	< 2	1	740	< 6	2	- 2
024	3	3	3	801	< 3	1	- 1				
025	5	4	- 4	80 $\bar{1}$	5	5	4	810	5	7	- 9
026	< 3	3	3	802	< 2	< 2	1	820	< 6	< 1	1
				803	2	2	2	830	< 6	2	- 3
031	15	15	-13	804	17	16	-16	840	< 6	1	2
032	< 2	1	- 1	805	6	5	- 5				
033	8	8	8	806	< 3	2	3	910	< 6	6	8
034	< 3	< 2	- 1	807	< 3	< 2	- 1	920	9	9	- 8
035	2	3	3					930	< 6	< 2	- 1
				1001	< 3	< 2	2	1010	< 6	3	4
041	5	4	- 5	100 $\bar{1}$	9	8	- 9	1020	< 6	< 1	- 1
042	< 2	< 2	1	1002	12	12	13				
043	5	5	4	1003	4	4	2	111	8	8	6
044	< 3	3	- 4	1004	< 3	< 2	- 1	11 $\bar{1}$	37	36	36
				1005	< 3	< 2	0	112	59	57	57
051	< 3	< 1	0	1006	6	6	- 7	113	< 2	2	5
052	4	4	4					223	3	3	- 3
053	< 3	< 1	- 1	1202	10	11	-11	222	13	12	14
				1203	9	9	- 8	221	8	7	5
206	< 3	4	3	1204	< 3	5	4	22 $\bar{1}$	7	7	8
205	< 3	2	2					222	9	9	- 7
204	17	17	17	110	15	15	13	223	21	19	19
203	13	13	12	120	4	4	- 1	224	< 3	2	4
202	15	16	-11	130	7	6	- 5	333	< 3	3	3
201	23	26	19	140	9	7	- 8	332	10	10	9
20 $\bar{1}$	13	15	12	150	< 6	2	- 2	331	8	7	7
202	70	72	91					33 $\bar{1}$	10	8	- 9
203	6	7	10	210	< 3	< 1	0	332	< 3	3	1
204	11	12	11	220	10	11	- 8	333	5	4	- 4
205	< 2	1	2	230	4	5	- 4	334	10	10	- 9
206	5	5	- 4	240	< 5	< 2	- 1	44 $\bar{1}$	8	6	7
207	< 3	< 2	0	250	10	8	9	442	< 3	< 2	- 2
								443	11	11	-10

regular hexagons of diameter 2.82 Å. Assuming that they are exactly so, their orientation to the crystal axes was calculated from their projected dimensions. The position of the CH_2 atom was located from the Fourier peaks and gave $\text{C}_6\text{H}_5\text{—CH}_2$ and $\text{CH}_2\text{—CH}_2$ distances of 1.47 and 1.58 Å respectively.

THE ADDITIONAL EXPERIMENTAL DATA

The crystallographic details were confirmed, the axis measurements agreeing with Robertson's within 0.01 Å.

The molecule of dibenzyl is essentially three-dimensional in that all the atoms do not lie in one plane, and in order to locate the atomic positions directly without any assumptions about the interatomic distances three-dimensional Fourier syntheses are necessary. All the X-ray reflexions within the range of $\text{Cu } K_\alpha$ radiation were therefore recorded using a single-crystal Unicam X-ray photogoniometer. The ranges of oscillation about the axes were chosen so that planes of low indices were generally recorded on ten different photographs, while other reflexions appeared less often according as the indices were beyond the limiting layer lines. In this way the stronger intensities were measured more accurately, congruous with attaching a greater significance to absolute rather than to proportional errors. To bring strong and weak reflexions from the same oscillation within the limits of accurate intensity measurement, a modified multiple-film technique (cf. Robertson 1943) was used, three films being interlayered with aluminium screens of absorption giving a correlating factor of four between each film. The intensities were estimated visually by matching with calibration spots made at the conclusion of each set of oscillations with the same batch of films and a suitable reflexion from the same axis setting. After correcting for the variation in angular velocity of oscillation, the mean values on an arbitrary scale were converted to absolute intensities by comparison with the ninety-six absolute intensities given by Robertson. The structure amplitudes for the 746 planes observed were then calculated with the usual formulae for Lorentz and polarization factors.

In table 1, Robertson's published values for the limited number of planes which he considered are compared with the new observations. In view of the different techniques employed, the agreement is very satisfactory and can be regarded as justification for the visual method of measuring intensities which, by reason of its rapidity of operation, becomes practically indispensable when a large number of experimental structure factors are required. The differences between the two results, including that for (200) which is probably an extinction effect, form an approximately normal error curve from which the most probable error is 0.6.

The approximate atomic parameters

That the structure proposed by Robertson was correct in all but the finer details was borne out by the agreement between his calculated and observed structure factors, and this was further confirmed by extending the comparison to many of

the new results. However, as was pointed out by Robertson (1939), with the limitations of the Fourier projection method, it was doubtful whether any significance could be attached to the difference between the observed acyclic bond lengths and the normal value of 1.54 Å, nor could it be said that the aromatic bond lengths were not 1.39 Å as determined in benzene itself by Brockway & Pauling (1934). A new set of co-ordinates was therefore derived for the purpose of commencing the Fourier refinement. Keeping the molecule with the same orientation to the crystal axes, the C—CH₂ and CH₂—CH₂ lengths were converted to 1.54 Å, and the benzene rings reduced to regular hexagons of diameter 2.78 Å. The actual changes in co-ordinates were small, the maximum shift was 0.11 Å and the mean 0.03 Å, and for structure factors calculated from this model the agreement with the experimental values was not appreciably reduced.

THE REFINEMENT OF THE ATOMIC PARAMETERS

All the new experimental results were included in the refining syntheses, but for the early stages, instead of using three-dimensional sections and lines for locating the maxima of the Fourier peaks, the less laborious section projections and projected sections of Booth (1945) were evaluated. A section projection on (001) with the limits $[c] = 0 \rightarrow \frac{1}{2}$ excluded the overlapping adjacent molecule in the cell and gave sharp and well-resolved peaks for the atoms in the benzene ring. For these special limits, the general formula for the synthesis reduces to

$$\begin{aligned} \rho(x, y)_{z=0}^z = \frac{1}{2\pi A} & \left\{ \sum \sum \sum_{h+k=2n} \pi F(h, k, 0) \cos 2\pi hx \cos 2\pi ky \right. \\ & + [F(h, k, l) - F(h, k, \bar{l})] \sin 2\pi hx \cos 2\pi ky \frac{(\cos \pi l - 1)}{l} \\ & + \sum \sum \sum_{h+k=2n+1} \pi F(h, k, 0) \sin 2\pi hx \sin 2\pi ky \\ & \left. + [F(h, k, l) - F(h, k, \bar{l})] \cos 2\pi hx \sin 2\pi ky \frac{(\cos \pi l - 1)}{l} \right\}. \end{aligned}$$

The C_1 peak from this synthesis is distorted owing to the limit $[c] = 0$ passing obliquely through the C'_1 - C_1 bond, and in figure 2, showing the contour map of the summation, the C_1 peak is obtained from a separate section at $z = 0.970$. Parallel to the (010) plane, two projected sections were computed, one containing C_1 , C_2 , C_4 and C_6 and the other C_3 , C_5 and C_7 . The position of these peaks determined the z co-ordinates and confirmed the x co-ordinates from the section projection. The new parameters gave rise to twenty sign changes on recalculation of the structure factors, and these were introduced by applying correcting syntheses. Further recalculation showed only six sign changes, and at this stage in the refinement three-dimensional sections at particular y values and lines parallel to the b axis were necessary to get the most accurate information available from the experimental data. These laborious summations were carried out by the Scientific Computing Service, as was also the

calculation of the final theoretical structure factors. In figure 3 the sectional contour maps are shown superimposed for the purpose of illustration on to the (010) plane. The contours are plotted at intervals of one electron per \AA^2 , the first being dotted

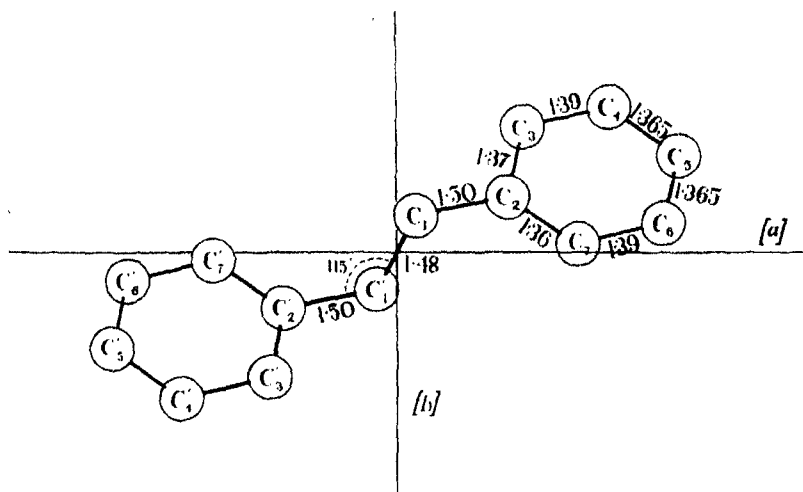


FIGURE 1

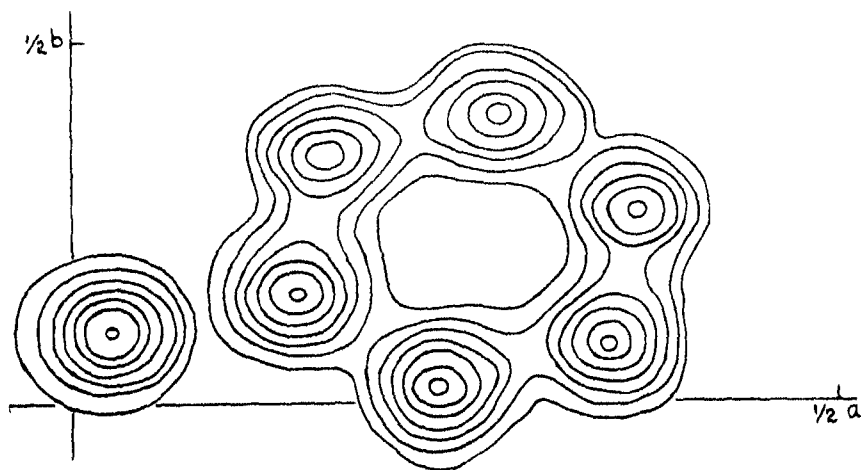


FIGURE 2

and the zero omitted for clarity in the diagram. There remained three sign changes for planes of very small amplitude which, when introduced as a correcting synthesis, produced a negligible shift on the co-ordinates (about 0.003 \AA).

The final atomic co-ordinates are given in table 2, and the corresponding intramolecular dimensions in table 3.

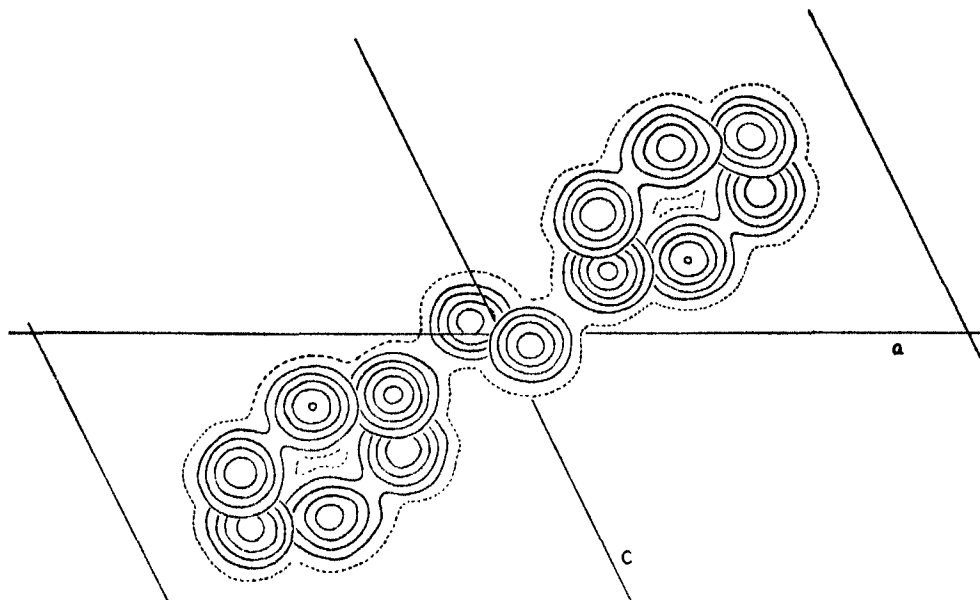


FIGURE 3

TABLE 2. ATOMIC CO-ORDINATES WITH RESPECT TO THE MONOCLINIC AXES

	<i>x</i>	<i>y</i>	<i>z</i>		<i>x</i>	<i>y</i>	<i>z</i>
C ₁	0.0260	0.096	-0.025	C ₆	0.3695	0.261	0.387
C ₂	0.1470	0.154	0.120	C ₆	0.3515	0.079	0.275
C ₃	0.1650	0.337	0.233	C ₇	0.2390	0.025	0.144
C ₄	0.2775	0.391	0.365				

TABLE 3. INTRAMOLECULAR DIMENSIONS (see figure 1)

C ₁ -C _{1'}	1.48 Å	C _{1'} C ₁ C ₂	115.0°	C ₅ -C ₆	1.365	C ₅ C ₆ C ₇	119.9°
C ₁ -C ₂	1.50	C ₁ C ₂ C ₃	120.0	C ₆ -C ₇	1.39	C ₆ C ₇ C ₈	120.0
C ₂ -C ₃	1.37	C ₂ C ₃ C ₄	119.6			C ₁ C ₆ C ₇	120.5
C ₃ -C ₄	1.39	C ₃ C ₄ C ₅	120.7	C ₇ -C ₂	1.36	C ₃ C ₄ C ₇	119.1
C ₄ -C ₅	1.365	C ₄ C ₅ C ₆	119.4				

Angle between plane containing C₂C_{1'}C₁C₂ and benzene rings - 71.6°.

C₁...C₃ = 3.48 Å, C₁'...C₇ = 3.18 Å.

The observed and calculated structure factors not given in table 1 are shown in an appendix. The agreement between the values is satisfactory, the ratio of $\sum |F_{\text{obs.}}| - |F_{\text{calc.}}|$ to $\sum |F_{\text{obs.}}|$ being 0.15. It is of interest to compare this with the closer agreement between the two sets of experimental values also given in table 1. It must be inferred that, owing to the assumption in the atomic scattering factors of isotropic thermal motion and spherically symmetrical electron distribution about the atomic co-ordinates, the structure factors have been measured more accurately

than they can be calculated. This does not, of course, affect the method of Fourier refinement which is independent of any assumptions about the form of the atomic scattering factors, but becomes important if the co-ordinates are derived by the method of direct adjustment to give the best agreement between observed and calculated values.

The difference between the observed and calculated structure factor for the very strong $20\bar{2}$ reflexion was unusually large, and, since it might represent an experimental error due to extinction, the effect of introducing a coefficient of +90 instead of +71 into the Fourier synthesis was considered. The corresponding shift in atomic co-ordinates was found to be less than 0.003 Å for any atom.

THE ACCURACY OF THE RESULTS

The atomic co-ordinates were obtained directly from the positions of the Fourier maxima without any reference to previous knowledge of the geometry of the benzene ring and its substituent groups. How exactly the results conform to what can reasonably be anticipated for a molecule of this type provides some basis for assessing their accuracy. In table 4, χ , ψ and ω are the inclinations to the $[a]$, $[b]$ and their perpendicular $[c']$ of the lines associated with the benzene rings.

TABLE 4

	χ	ψ	ω		χ	ψ	ω
C ₁ -C ₂	45.4°	76.3°	47.9°	C ₂ -C ₆	83.9°	35.6°	55.1°
C ₂ -C ₅	45.2	76.4	47.8	C ₂ -C ₇	36.5	54.5	82.7
C ₃ -C ₄	44.4	76.1	49.0	C ₄ -C ₈	36.6	54.1	82.4
C ₆ -C ₇	44.3	76.0	49.1	C ₃ -C ₇	121.7	35.9	75.0
C ₂ -C ₃	84.0	35.4	55.2	C ₄ -C ₆	121.7	35.9	75.0

It will be noted that within a fraction of a degree C₁-C₂ is collinear with C₂-C₅, and opposite sides of the benzene ring are parallel. On the other hand, there is a discrepancy of 1° between C₁-C₂-C₅ and C₃-C₄, C₆-C₇. This corresponds approximately to the difference of 0.01 Å between the bond lengths C₂-C₅ and C₂-C₇ and is certainly experimental error. All the angles associated with the benzene rings are 120° within $\pm 1^\circ$ and the variations inside that limit cannot be regarded as significant. The deviation of the C₁' C₁ C₄ angle from the normal tetrahedral angle of 109.5° is, however, quite definite and is one of the special features of the molecular structure. In testing for planarity of the ring it was found that both the planes C₂ C₃ C₇ and C₂ C₄ C₈ were inclined at exactly 71.6° to C₁' C₁ C₂, but the atoms C₃, C₅ and C₇ were each 0.02 Å out of the plane of C₂ C₄ C₆. The symmetry with which the benzene bond lengths fall into two classes according to their position in relation to the substituent bond is evidence for the acceptance of the difference between 1.37 and 1.39 Å as real, but whether the C₁'-C₁ and C₁-C₂ difference of 0.02 Å is also significant is less certain.

Summarizing, it appears from the self-consistency of the molecular dimensions that the bond lengths can be regarded as reliable to ± 0.01 Å and the angles to $\pm 1^\circ$.

Since many of the detailed structure analyses have in the past only proceeded as far as Fourier projection methods for determining the molecular dimensions, it is of interest from the point of view of assessing their accuracy to compare the refined atomic positions with those derived by Robertson. Bearing in mind, however, that dibenzyl, being a non-planar molecule of some complexity, was difficult to define from the two useful projections on (010) and (001), in neither of which were all the atoms completely resolved. The general orientation of the molecule in the unit cell was described quite accurately. Robertson's values corresponding to those in table 4 being

C_2-C_8	χ	46.7°	C_3-C_7	χ	120.6°
	ψ	77.3		ψ	34.3
	ω	46.1		ω	76.0

It was in the individual atomic co-ordinates that considerable shifts were found. The largest change and the mean change for the co-ordinates referred to the three crystal axes are shown below expressed in Angstrom units:

	x	y	z
maximum change	0.02	0.10	0.08
average change	0.01	0.04	0.05

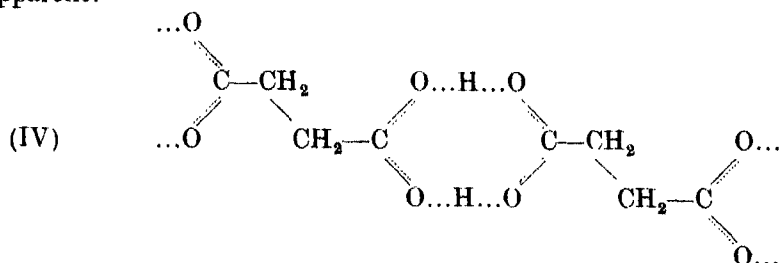
The accuracy of Robertson's x co-ordinates is most striking and shows that under the best conditions reliable parameters can be obtained from Fourier projections. The y and z values were each derived from a single projection, there being no resolution of the peaks in the [100] direction, and as might be expected the biggest error is from the poorest projection, that on (001).

DISCUSSION OF THE MOLECULAR STRUCTURE

The three acyclic bonds in the dibenzyl molecule are shorter than the normal C-C bond length by amounts which can confidently be regarded as well beyond the probabilities of experimental error. Together with the similar results from the crystal structure analysis of geranylamine hydrochloride by Jeffrey (1945), this is evidence of the existence of special structural properties associated with the system of two unsaturated groups separated by three single-carbon bonds. In dibenzyl the three bond lengths in question are 1.50, 1.48 and 1.50 ± 0.01 Å, whereas in the di-isoprene compound they were 1.51, 1.44 and 1.51 ± 0.04 Å. The estimated limits of experimental error in the latter analysis were very much greater than in the present work owing to difficulties inherent to the more complex molecule, and they do in fact include the variation between the two results. Whether the bond lengths are truly identical is uncertain at the moment. They might differ for two reasons: (a) that the electronic character of the substituted atom in the benzene ring is not the same as a doubly bound carbon atom although they are said to behave similarly in respect to first-order conjugation, comparing, for example, hexatriene and stilbene; (b) that an effect characteristic of the hydrocarbon is further enhanced in the geranylamine

molecule by the terminal NH_3^+ group as tentatively suggested by Bateman & Jeffrey (1943). It is anticipated that this question will be answered both by increasing the accuracy of the atomic co-ordinates from the geranylamine structure, using the method proposed by Booth (1945), and by undertaking other detailed structure analyses on appropriate hexadiene derivatives.

The third compound for which structural data are available at present is succinic acid. In the crystalline state, depicted as in (IV), a resemblance to dibenzyl is apparent:



The C-C bond lengths reported from crystal structure analysis by Verweel & MacGillavry (1939) are 1.50, 1.52 and 1.50 Å, which the authors conclude are not significantly different from normal. This implies an experimental error of at least 0.04 Å, and it may well be a little greater, since the atomic parameters were derived from two Fourier projections in one of which the resolution and shape of the peaks were poor. While it is of interest to note that short values were obtained for all three bonds, the results are of little significance to the problem without further refinement.

In all three molecules, the configuration of the adjacent carbon bonds about the central $\text{CH}_2\text{-CH}_2$ bond is planar, although this plane is not necessarily the same as or perpendicular to that of the benzene rings or the double-bond valencies. With the dibenzyl and succinic acid structures this is a condition determined by the crystal symmetry, but not with the geranylamine molecule where it only emerged as a result from the detailed analysis. Since there is no reason why the rotation should be rigidly restricted about a normal single bond in the central position of these molecules, it is suggestive that a planar configuration is an accompanying feature to the peculiar bond lengths.

As was pointed out when describing the di-isoprene structure a shortened central C-C bond in the 1.5 diene system cannot be simply explained in terms of the present theory of quantum mechanical resonance, and neither can the more exact picture revealed by this structure analysis. The reduction from the normal single-bond lengths may be regarded as due to the partial non-localization of the electrons of the C-H bonds and their transference to the carbon chain, designated by Bateman & Jeffrey (1945) as chain-hyperconjugation in Part IV of this series. From an alternate point of view in terms of atomic orbitals, it may be considered that the s/p hybridization ratio is modified so as to increase at the expense of the C-H bonds the s character of the C-C bonds as compared with a normal single sp^3 bond, assigning to the new mode of bond formation the observed interatomic distances. With either

point of view it must be postulated that the molecule gains extra 'resonance' energy when instead of complete insulation of the unsaturated groups by the three C-C links some degree of electronic interaction is possible. If analogy with the series butadiene, hexatriene, octatetrene is permissible, then this effect will be intensified with increase of the number of 1.5 diene units, and in the polyisoprenes, rubber, gutta-percha and chloroprene, a structural modification of the bond characters as great if not greater may be anticipated.

In the dibenzyl molecule the shortening of the bonds between the two benzene rings appears to be reflected by a redistribution of the bond lengths within the rings. While two of the aromatic bonds have the value 1.39 Å, which is that generally accepted for benzene itself, the other four are shorter with 1.37 Å. No comparable experimental data for fully conjugated molecules are at present available, since in the X-ray analysis of stilbene, azobenzene, tolan and diphenyldiacetylene it is assumed in order to determine the orientation of the molecules that the benzene ring is a perfect hexagon of sides 1.39 Å. The theoretical calculations of Penney & Kynch (1938) on phenylethylene for the bonds here referred to as C_2-C_3 , C_3-C_4 , C_4-C_5 were 1.398, 1.390 and 1.393 by the approximation of electron pairs and 1.374, 1.377 and 1.374 by the orbital method. Unfortunately, the limits of experimental error in the present work are too large to be able to define as precisely as would be desired the differentiation between the two types of aromatic bond, but the results do indicate, first, that the difference in length may be somewhat greater than theoretically predicted even for a fully conjugated system, and secondly, that either the benzene ring as a whole acquires a greater degree of double-bond character in this molecule or the value of 1.39 Å for benzene itself is about 0.01 Å too large. The most accurate experimental derivations of this standard value are reported from electron diffraction on benzene by Brockway & Pauling (1934), 1.39 ± 0.01 Å, and from X-ray diffraction on hexamethyl benzene by Brockway & Robertson (1939), 1.39 ± 0.02 Å. The theoretical treatment of Lennard-Jones & Turkevitch (1937) by the method of molecular orbitals gave a value of 1.37 Å, starting from the assumption of 1.54 and 1.33 Å for the pure single- and double-bond lengths.

Of interest as an example of the directive influence that intermolecular packing can have on the intramolecular configuration is the orientation of the CH_2-CH_2 bond at $70\frac{1}{2}^\circ$ to the plane of the benzene rings. In a comparatively isolated state of gas or liquid phase the molecule would be expected to be symmetrical with this central bond perpendicular to the benzene rings and $(CH_2)\dots C_3$ equal to $(CH_2)\dots C_7$. In the crystal these two distances differ by 0.3 Å, and the small unbalance between the Van der Waals forces must be compensated by a more compact packing than could be attained by absolutely symmetrical molecules. The intermolecular distances are very similar to those described by Robertson in the original paper, the closest approaches of 3.7–4.1 Å being between adjacent benzene rings.

This work forms a part of the programme of fundamental research on rubber undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

- Bateman & Jeffrey 1943 *Nature*, **152**, 446.
 Bateman & Jeffrey 1945 *J. Chem. Soc.* p. 211.
 Bateman & Koch 1944 *J. Chem. Soc.* p. 600.
 Bauer & Beach 1942a *Amer. Min.* **27**, 214.
 Bauer & Beach 1942b *J. Amer. Chem. Soc.* **64**, 1142.
 Booth 1945 *Trans. Faraday Soc.* **41**, 434.
 Booth 1945 *Nature*, **156**, 51.
 Brockway & Pauling 1934 *J. Chem. Phys.* **2**, 867.
 Brockway & Robertson 1939 *J. Chem. Soc.* p. 1324.
 Jeffrey 1945 *Proc. Roy. Soc. A*, **183**, 388.
 Lennard-Jones & Turkevitch 1937 *Proc. Roy. Soc. A*, **158**, 297.
 Penney & Kynch 1938 *Proc. Roy. Soc. A*, **164**, 409.
 Robertson 1934 *Proc. Roy. Soc. A*, **146**, 473.
 Robertson 1935 *Proc. Roy. Soc. A*, **150**, 348.
 Robertson 1939 *Rep. Progr. Phys.* p. 347.
 Robertson 1943 *J. Sci. Instrum.* **20**, 175.
 Verweel & MacGillavry 1939 *Z. Krist.* **102**, 60.

APPENDIX

Observed and calculated structure factors

(Additional to those given in Table 1)

<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}
007	2	2	115	4	4	313	6	9
008	4	4	116	8	9	314	30	28
208	5	— 5	117	< 2	0	315	2	3
406	3	4	114	6	— 5	316	< 2	— 1
408	< 2	— 1	115	< 1	0	317	< 2	— 1
605	< 2	0	116	6	— 6	318	2	— 2
608	3	4	117	< 2	0	411	25	22
804	< 2	— 1	118	2	2	412	3	— 4
808	2	— 3	211	9	— 7	413	4	— 5
1002	< 2	— 1	212	5	2	414	1	1
1003	2	2	213	15	14	415	< 2	1
1007	2	— 2	214	2	— 3	416	< 2	— 1
1008	3	— 3	215	2	2	411	21	17
1200	< 2	0	216	< 2	2	412	13	13
1201	< 1	1	211	54	63	413	21	21
1201	1	3	212	8	— 5	414	2	— 2
1205	< 2	— 1	213	16	— 15	415	6	6
1206	< 2	1	214	9	9	416	3	3
1207	< 2	0	215	6	— 4	417	3	— 3
1402	< 2	0	216	1	3	418	< 2	1
1403	< 2	0	217	< 2	0	511	7	7
1404	3	— 4	218	< 2	1	512	2	— 2
1405	3	— 2	311	7	5	513	3	4
1406	2	— 1	312	12	10	514	4	3
			313	9	9	515	3	2
017	5	5	314	9	9	511	2	— 2
018	< 1	0	315	< 2	1	512	15	18
112	4	— 5	316	< 2	— 2	513	5	5
113	1	4	311	10	9	514	17	— 16
114	5	— 3	312	39	— 34	515	2	— 2

*Observed and calculated structure factors (continued)**(Additional to those given in Table 1)*

<i>hkl</i>	<i>F</i> _{obs}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}
516	3	3	1013	< 2	- 2	324	2	3
517	< 2	1	1014	2	3	325	4	- 4
518	< 2	0	1015	3	- 3	326	2	0
611	6	- 6	1016	< 2	1	321	3	- 6
612	3	- 4	1017	< 2	0	322	10	10
613	2	0	1018	< 2	- 1	323	33	38
614	< 2	0	1110	3	- 3	324	< 2	1
615	2	- 3	1111	< 2	1	325	9	- 8
611	18	- 16	111	6	- 5	326	< 2	2
612	5	- 2	1112	4	- 3	327	< 2	1
613	16	16	1113	< 2	- 1	421	2	- 2
614	7	7	1114	3	3	422	3	3
615	1	2	1115	< 2	0	423	2	2
616	< 2	1	1116	< 2	0	424	7	- 6
617	5	5	1117	< 2	0	425	< 2	1
618	< 2	1	1118	4	- 4	421	9	10
711	9	8	1210	< 2	- 1	422	< 2	2
712	9	10	1211	1	- 2	423	12	- 10
713	< 2	- 1	1212	3	- 3	424	15	- 13
714	< 2	- 1	1213	7	7	425	6	8
711	< 2	2	1214	2	2	426	< 2	1
712	< 2	1	1215	< 2	0	427	< 2	- 1
713	7	- 6	1216	< 2	1	521	3	- 5
714	< 2	1	1311	< 2	0	522	2	2
715	< 2	- 1	1312	< 2	0	523	< 2	- 1
716	< 2	0	1313	3	- 1	524	4	- 4
717	< 2	- 1	1314	5	- 6	525	< 2	- 1
718	2	- 2	1315	2	- 3	521	5	6
811	7	- 7	1316	< 2	1	522	14	14
812	1	1				523	3	- 4
813	1	- 3	027	2	2	524	10	11
814	< 2	0	121	10	- 13	525	17	16
811	2	1	122	3	- 4	526	8	8
812	4	5	123	< 2	0	527	< 2	- 1
813	8	- 8	124	4	3	621	2	- 2
814	2	- 1	125	12	12	622	3	- 4
815	6	7	126	< 2	2	623	6	6
816	< 2	0	127	2	- 2	624	3	3
817	< 2	1	121	25	29	621	5	- 3
818	1	1	122	5	3	622	14	14
911	5	- 5	123	20	- 18	623	1	2
912	4	- 4	124	2	3	624	7	8
913	< 2	1	125	< 2	0	625	< 2	- 1
911	5	5	126	3	3	626	8	- 7
912	2	- 1	127	7	6	627	< 2	1
913	2	0	128	1	1	721	1	4
914	< 2	- 2	224	7	8	722	8	- 9
915	2	- 2	225	3	4	723	6	- 6
916	7	- 6	226	4	- 3	724	< 2	1
917	2	- 3	225	2	- 2	721	10	11
918	1	2	226	< 2	1	722	2	- 2
1011	< 2	0	227	2	- 2	723	< 2	2
1012	< 2	- 1	321	5	- 4	724	1	2
1011	10	11	322	4	4	725	4	4
1012	7	6	323	9	9	726	6	6

<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}
727	2	2	131	10	10	733	4	4
728	< 2	1	132	8	- 8	734	< 2	1
821	8	7	133	7	- 5	731	4	- 5
822	< 2	1	134	< 2	1	732	6	- 4
823	2	- 3	135	< 2	1	733	2	- 1
821	1	- 1	136	2	2	734	6	6
822	10	- 10	137	< 2	- 1	735	8	8
823	3	- 2	231	9	10	736	< 2	1
824	4	3	232	8	- 8	737	3	- 3
825	4	- 4	233	4	4	831	2	- 1
826	4	4	234	6	7	832	6	- 5
827	2	2	235	3	- 4	833	< 2	- 1
828	< 2	- 1	236	< 2	1	831	< 2	0
921	7	- 8	231	14	14	832	4	2
922	2	2	232	8	10	833	5	7
923	< 2	1	233	3	4	834	4	- 4
921	< 1	- 1	234	2	- 3	835	3	5
922	3	5	235	< 2	4	836	4	5
923	5	5	236	< 2	2	837	2	- 3
924	< 2	1	237	< 2	0	931	< 2	1
925	5	- 3	334	7	- 6	932	< 2	1
926	3	- 2	335	5	- 5	931	4	3
927	4	5	335	< 2	1	932	2	2
1021	2	- 3	336	2	3	933	4	- 3
1022	< 2	1	337	< 2	2	934	6	- 4
1021	2	2	431	2	- 2	935	3	- 3
1022	< 2	1	432	5	6	936	3	3
1023	2	2	433	9	- 10	937	3	3
1024	2	- 3	434	2	1	1030	3	- 2
1025	< 2	0	435	4	3	1031	< 2	0
1026	< 2	- 1	431	< 2	- 2	1031	5	5
1027	4	- 4	432	10	10	1032	3	4
1120	< 2	0	433	< 2	- 1	1033	2	- 2
1121	< 2	1	434	7	9	1034	< 2	2
1122	< 2	- 1	435	9	9	1035	< 2	3
1123	3	- 2	436	< 2	- 3	1036	2	- 3
1124	4	- 4	531	1	- 1	1037	< 2	1
1125	< 2	- 1	532	8	- 7	1132	< 2	- 1
1126	3	- 2	533	7	- 7	1133	< 2	3
1220	2	1	534	< 2	0	1134	< 2	1
1221	< 2	- 1	531	5	5	1135	< 2	0
1222	< 2	0	532	5	5	1230	2	- 1
1223	2	- 2	533	5	6	1231	< 2	- 1
1224	2	- 2	534	1	1	1232	< 2	0
1225	< 2	1	535	9	- 8	1233	< 2	2
1323	< 2	- 1	536	7	- 5			
1324	< 2	- 1	631	7	- 7			
1325	4	4	632	< 2	0	045	< 2	1
1326	2	1	633	3	3	141	2	2
			634	4	- 4	142	6	7
036	4	4	631	8	9	143	4	3
037	< 2	1	632	< 2	0	141	3	4
131	16	- 16	633	5	6	142	4	- 5
132	2	- 2	634	9	10	143	2	2
133	5	3	635	6	- 6	144	6	7
134	7	6	636	3	4	145	< 2	1
135	< 2	1	731	4	- 3	241	1	- 3
136	3	- 3	732	< 2	0	242	< 2	- 3

Observed and calculated structure factors
(Additional to those given in Table 1)

<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}	<i>hkl</i>	<i>F</i> _{obs.}	<i>F</i> _{calc.}
243	5	5	843	4	- 3	751	< 2	- 1
244	< 2	- 3	940	2	3	751	4	- 5
245	< 2	- 1				752	< 2	- 2
			054	< 2	0	753	3	3
241	12	- 13	151	4	- 4	850	3	4
242	5	- 5	152	3	- 3	851	3	- 3
243	< 2	- 1	153	5	- 5	852	3	4
244	< 2	1	151	1	- 2	853	< 2	1
245	< 2	1	152	< 2	- 2			
246	2	- 3	153	< 2	2	061	2	3
341	< 2	1	154	< 2	- 2	062	2	- 2
342	< 2	0	155	5	- 5	063	5	- 5
343	5	- 4	251	< 1	0	160	2	- 2
344	< 2	- 2	252	< 2	- 1	161	2	1
345	< 2	0	253	< 2	0	162	2	1
341	3	- 4	254	< 2	- 1	163	< 2	0
342	2	1	255	< 2	0	161	2	0
343	4	5	251	< 1	0	162	< 2	- 1
344	2	1	252	5	5	163	< 2	0
345	< 2	3	253	2	1	164	2	- 1
346	3	3	254	< 2	1	260	2	- 2
441	5	5	255	< 2	- 1	261	7	- 8
442	6	- 7	351	2	1	262	< 2	0
443	6	- 6	352	2	- 3	263	< 2	1
444	< 2	1	353	< 2	- 1	261	2	2
445	5	- 3	354	< 2	- 1	262	< 2	- 1
444	< 2	1	351	6	- 7	263	1	1
445	2	- 3	352	< 2	0	264	< 2	- 1
446	< 2	0	353	2	- 3	360	3	3
541	7	- 8	354	< 2	- 1	361	< 2	0
542	4	- 4	355	2	3	362	< 2	0
543	3	3	451	3	- 2	363	< 2	0
544	2	3	452	3	- 4	361	< 2	- 1
541	4	3	453	< 2	0	362	< 2	0
542	7	8	451	< 2	0	363	< 2	0
543	3	- 3	452	10	- 10	364	< 2	1
544	4	- 4	453	2	- 1	460	< 2	0
545	< 2	1	454	2	3	461	< 2	1
546	3	3	455	< 2	1	461	5	- 6
641	9	- 9	551	5	- 5	462	< 2	- 1
642	2	2	552	< 2	0	463	< 2	1
643	< 2	- 1	553	< 2	- 2	464	< 2	0
641	< 2	0	551	5	6	560	2	- 2
642	< 2	0	552	< 2	- 2	561	< 2	0
643	10	10	553	2	- 3	562	< 2	- 2
644	< 2	- 1	554	< 2	1	563	< 2	- 1
741	2	3	555	< 2	- 1	660	< 1	0
741	3	5	650	4	- 5	661	3	3
742	< 2	0	651	2	2			
743	5	5	652	2	3	071	< 2	0
744	6	6	651	3	2	072	3	- 3
841	< 2	2	652	3	4	270	2	- 2
841	3	5	653	2	1	271	2	- 1
842	3	- 1	654	5	- 5	272	3	1
			750	2	2	470	2	2

The internal ballistics of a leaking gun

By J. CORNER, *Armament Research Department, Ministry of Supply*

(Communicated by N. F. Mott, F.R.S.—Received 4 April 1946)

[Plate 3]

The equations of internal ballistics are developed for a gun in which the propellant gases can escape by a vent, whose area can be, in the most general case, a function of time or travel. The most important special cases are treated in detail. The theory has been applied to the testing and design of recoilless guns, to the leakage past the projectile in smooth-bore mortars and to the calculation of the loss of performance as a gun wears.

1. INTRODUCTION

Though the literature on internal ballistics is voluminous (cf. Cranz 1926), it has been impossible to find any previous work on the problem treated here, where the gas can escape to the atmosphere. This is all the more surprising when the many possible applications are realized. The most spectacular of these is the recoilless gun, which discharges gas through a venturi in the breech in order to cancel the recoil (figure 1, plate 3 and figure 2). It was, in fact, the appearance of such weapons which led Colonel G. H. Hinds to suggest to the writer that this subject was worthy of study. The methods developed for this problem have proved to have applications also to smooth-bore guns, and even to completely orthodox guns which have become worn.

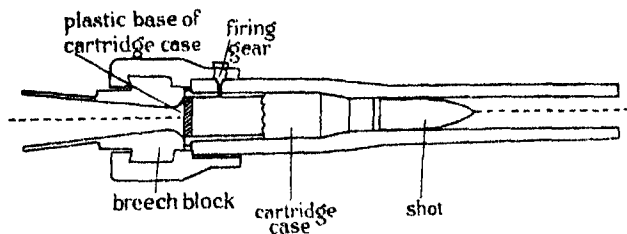


FIGURE 2. A typical recoilless gun, in section, showing the plastic disk which provides an initial seal to the venturi.

It is convenient to mention at this point some of the technical matters to which the writer and his collaborators have applied the equations of this paper. The results themselves are mainly of engineering interest, and are, in any case, confidential. Problems treated in connexion with the recoilless gun have included the optimum design of such weapons, their best field of application, the possibilities of using a muzzle brake on a recoilless gun, as suggested by Cooke (1921), the best bursting pressure for the disk which initially seals the nozzle, adjustments required for

extreme climatic conditions, and the examination of certain claims made for such guns. The gas leak past the projectile in a smooth-bore gun is another field of application for our methods. In the more elaborate of such guns this leak is prevented by an obturating device on the shot, but this is hardly possible on the muzzle-loading smooth-bore mortar. This is, next to the rifle, the most numerous of all weapons, and a good case can be presented for the thesis that the mortar is the most important of all. The clearance between bomb and barrel causes the muzzle velocity to drop by about 5 %. This would not be important were it not that bombs differ in diameter. The resulting spread in velocities, and hence in range, was calculated from this theory, and later confirmed by check firings. It was found that firings carried out on three types of mortar in two countries, analysed by independent statisticians, agreed with the theory.

In a worn orthodox gun, firing shot which are fixed in the mouth of a cartridge case, there is a considerable 'run-up' before the driving band of the shot engages the rifling. This run-up may amount to as much as several calibres. There is, consequently, a loss of gas in the earliest stages of the firing, and this is responsible for some of the loss of performance as a gun wears. This application differs from those mentioned earlier in that the leakage area for the escape of gas is not constant, being now a rapidly varying function of shot travel. The equations of this paper are still applicable, though not the simplest of the methods of solution. The work in this field has shown that the worn gun can be treated in a manner as exact as that in which orthodox internal ballistics deals with a new gun.

2. OUTLINE OF THE THEORY

The assumptions underlying our more accurate method (§ 5.1) are listed, starting with those points which occur also in the internal ballistics of orthodox guns. This will make clear the relation to normal internal ballistics.

Assumptions

- (i) The propellant burns in parallel layers.
- (ii) The covolume per unit mass of propellant gases is independent of their temperature. This is a normal and accurate assumption.

The following assumptions are associated with the features peculiar to the vented gun:

- (iii) No unburnt cordite is lost through the nozzle. This is obviously true for smooth-bore guns with a small leakage area between shot and bore, and also for orthodox worn guns. In many types of recoilless guns the cordite is effectively trapped by the cartridge case. Even in the most unfavourable case, that of guns such as shown in figure 2, the losses of unburnt cordite are small unless the charge burns very slowly.

- (iv) It is assumed that the bursting of the disk which closes the nozzle and the setting up of the flow out of the gun can be represented by the use of the familiar

equations for quasi-steady flow through nozzles, beginning instantaneously at a certain pressure. This will be called the 'nozzle-start pressure'.

The use of this idealization requires a certain amount of care. For guns with thin bursting disks, such as the paper or sheet brass used in certain designs, the breaking of the disk occurs at a time when the rate of increase of pressure is small; hence, in the time taken by the flow through the nozzle to start and settle down, the pressure in the chamber does not rise much. The 'nozzle-start pressure' is low—indeed, such guns can usually be treated on the assumption of a zero nozzle-start pressure. Thickening the disk delays the initial nozzle flow until an epoch where the pressure is rising rapidly, and the 'nozzle-start pressure' under these conditions is considerably greater than that needed to break the disk. The apparent 'nozzle-start pressure' would therefore be expected to show a very rapid increase with disk thickness once this has passed a certain critical value. This appears to be borne out by experience with German recoilless guns, where plastic disks as thick as a third of a calibre have been used.

In the basic equations and in the more exact of the two methods of computation derived from them, it is *not* assumed that the nozzle throat area is constant during the firing. In most recoilless guns the throat area is constant; in certain other cases of vented weapons, however, the vent area varies during the firing. This is the case, for example, with the leakage in worn orthodox guns. Such problems can be attacked by the method of numerical integration given in § 5.1.

All covolume corrections are included, apart from those which can be proved to be trivial.

Corrections for the pressure distribution down the bore of the gun are quite different in the recoilless gun from those in the normal gun. In the latter there is a stagnation point at the breech; in recoilless guns the stagnation point is not fixed, and spends most of its time near the front of the chamber or in the bore. The ordinary Lagrange corrections cannot be used in the recoilless gun. Lagrange corrections have been omitted, and therefore the equations are not applicable to guns of high muzzle velocity.

It is assumed that the rate of burning is proportional to some power of the pressure. This is the conventional assumption where a form of a wide generality is required. However, there is no difficulty in writing down the equations with any other law of burning.

Now consider the more rapid analytical method, given in § 5.2. This uses the further assumption that the rate of burning is proportional to pressure.

In this method the interaction of the shot with the gun as an idealized engraving resistance which disappears as soon as the projectile moves. This is no doubt a violent approximation, but in orthodox ballistics there are at least three justifications. This behaviour is easily included in the ballistic equations; physically more accurate assumptions would lead to more parameters, which it would be difficult to evaluate experimentally; in any case the calculated ballistics are not sensitive to the way in which the resistance to motion is taken into account; for example, the

replacement of a shot-start resistance by a 'modified web size', chosen to give the same maximum pressure, has very little effect on the muzzle velocity and pressure-space curve.

In normal practice the shot-start pressure is chosen to reproduce the observed maximum pressure in the gun, and the overall accuracy of the calculation is checked by the comparison of muzzle velocities. Any factor which has been omitted from the ballistic calculations and which affects the maximum pressure, appears in the 'shot-start pressure'. For example, the heat loss to the bore is often neglected. This means that the 'shot-start pressure' necessary to give the observed maximum pressure then depends on the heat loss to the walls as well as on the nature of the band and rifling. This well-known fact appears in an obvious form in experiments with unorthodox bands. However, the muzzle velocity is not sensitive to such imperfections of the 'shot-start' model and usually good agreement can be obtained.

In the more accurate work (§ 5.1) there is no restriction to this particular idealization of the initial resistance to motion, and in particular cases resistances which were functions of the shot travel have been used, sometimes in combination with a shot-start pressure.

3. THE EQUATIONS OF INTERNAL BALLISTICS OF A LEAKING GUN

(3.1) *Notation*

Let the area of the cross-section of the bore be A , and the chamber volume, measured out to throat of nozzle, be U . Let the charge weight be C and shot weight be W . Take l as the total shot travel and S as the area of the throat through which the gas escapes. Write the shot travel at time t as x and the velocity as V . Let the fraction of charge burnt be ϕ , and the mass of gas present in the gun be NC . Denote the remaining least dimension of propellant grain by fD , where D is the initial web size of the cordite. The quantities ϕ and f are connected by the geometry of the partially burnt pieces of propellant, a relation which in British practice is usually written as

$$\phi = (1-f)(1+\theta f), \quad (1)$$

where θ is the 'form-factor' of the grain. For example, θ is zero for long tubes and unity for long cords. Not all possible forms of the relation between ϕ and f can be represented by (1); propellant chopped into short lengths shows higher terms in f . However, all propellant shapes found in practice either obey (1) exactly or can be approximated satisfactorily by a suitable constant θ . (1) will be used in the remainder of this paper.

Let the propellant have covolume η , density δ , diametral rate of burning βP^a at pressure P , and let the uncooled explosion temperature (without performance of external work) be $T_0^\circ \text{K}$. Let R be the gas constant per g.mol., divided by the mean molecular weight of the propellant gases, so that RT_0 is the quantity usually known

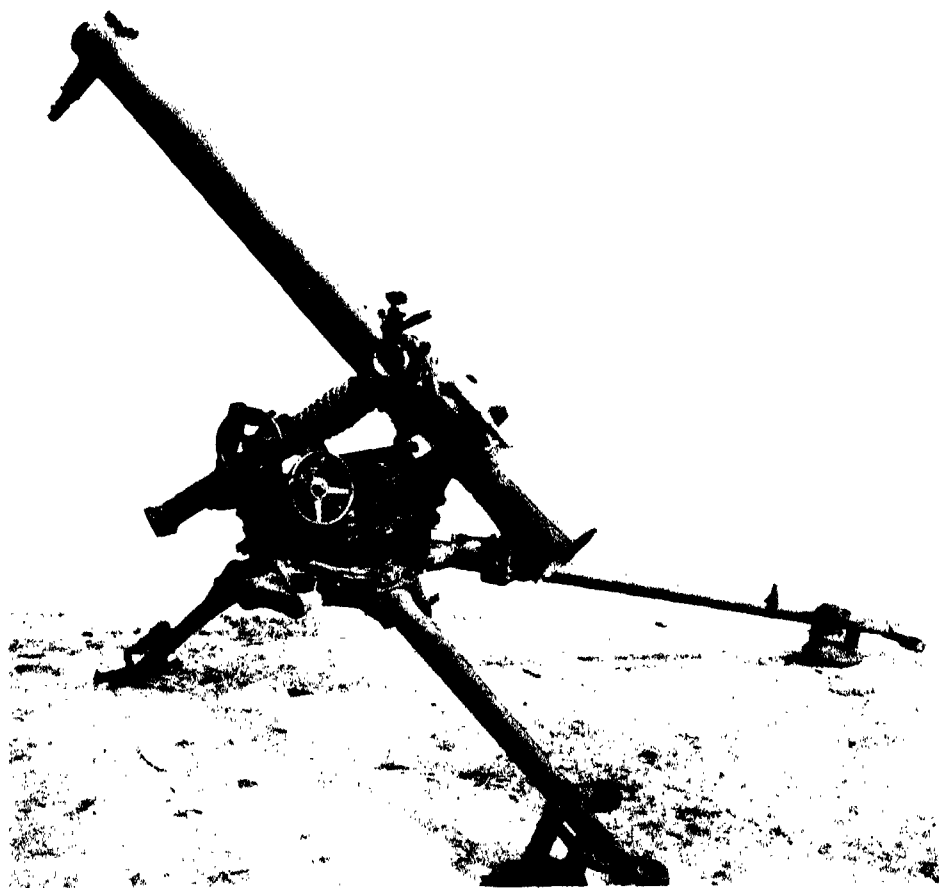


FIGURE 1. A German recoilless gun of 15 cm. calibre, which fired an 84 lb. projectile at 1020 ft./sec. Weight of equipment, as shown, is 1400 lb.

as the 'force constant' of the propellant. The equation of state of the products of uncooled explosion is

$$P(v - \eta) = RT_0, \quad (2)$$

where v is the volume per gram.

(3.2) Pressure and density in the gun

The standard theory assumes that the density of gas is independent of position in the gun. This assumption is used here; let the density be ρ . Hence

$$\rho \left[U + Ax - \frac{C}{\delta} (1 - \phi) \right] = NC. \quad (3)$$

Let P be the mean pressure in the gun, and $T^\circ \text{K}$ the mean temperature of the gas in the gun, both at time t . As corrections for pressure gradient are omitted, P is taken also as the pressure on the shot and as the mean pressure acting on the burning propellant.

Therefore,

$$P \left(\frac{1}{\rho} - \eta \right) = RT,$$

which from (3) becomes

$$P \left[U + Ax - \frac{C}{\delta} (1 - \phi) - CN\eta \right] = NCRT. \quad (4)$$

Numerical examples have shown that for most guns with a substantial leak, such as recoilless guns with nearly balanced recoil, this equation may be replaced by

$$P \left(U + Ax - \frac{C}{\delta} \right) = NCRT, \quad (5)$$

with ample accuracy, right up to the time when the charge is all burnt, $\phi = 1$. In a typical case, the error in the pressure due to the use of (5) was at most 0.5 %, which occurred in the early stages; the error decreased with time until $\phi = 1$, and was 0.3 % when $\phi = \frac{1}{2}$. This is about a tenth of the corresponding error in an orthodox gun at the same stage of burning.

In exceptional cases errors have been found of as much as 10 % in pressure due to this approximation. It has been found that for a given gun and projectile these percentage errors are almost independent of the loading and initial conditions, and are substantially constant over the important period of high pressures in the firing. These observations justify the following method of correction: equation (5) is always used, and on first calculations with a new gun the percentage error at maximum pressure is found due to the use of (5) instead of (4). If this error is appreciable all future runs are corrected by this factor. In the exact method it is necessary only to multiply $NCRT$ by the appropriate factor to get the corrected pressure. In the approximate method (§ 5.2) the correction is absorbed into RT_0 , using the corrected value, say RT_{01} , to replace RT_0 only in equations (40), (50), (51), (56) and (57).

Where the leakage is small, as in smooth-bore guns and eroded guns, the approximation (5) is as inaccurate as in orthodox internal ballistics.

After $\phi = 1$, the approximate equation (5) becomes inadequate in all cases, and in this part of the solution the exact equation (4) will be used.

The equation of motion of the shot is

$$W \frac{d^2x}{dt^2} = AP. \quad (6)$$

The law of burning is

$$D \frac{df}{dt} = -\beta P^\alpha, \quad (7)$$

and, as already explained, the connexion between ϕ and f is

$$\phi = (1-f)(1+\theta f). \quad (8)$$

It should be mentioned that the geometrical interpretation of the variable f and of the relation between ϕ and f is obscured by inhomogeneity of the charge, imperfections of charge units, failure to burn by parallel layers, and certain other effects. Nevertheless, it is still mathematically convenient to use (7) and (8), with f merely a certain auxiliary variable which could be eliminated to give a single, more complicated, equation expressing the burning law.

(3.3) Flow through the nozzle

Let $C_p/C_v = \gamma$ for the propellant gases at the temperatures in the gun. It is convenient to introduce a quantity ϵ , defined by

$$\epsilon = \eta\rho/(1-\eta\rho). \quad (9)$$

For normal propellants ϵ is rarely more than 0.35, this corresponding to pressures in the neighbourhood of 25 tons/sq.in.

The gases leave the chamber and enter the nozzle with an initial velocity. This can be taken into account by using a fictitious pressure and temperature (in the usual equations for flow out of a reservoir), such that the gases could expand from rest at this pressure P_r and temperature T_r down to the velocity, pressure and temperature which they possess at the nozzle entry. In the present work the reservoir pressure and temperature are taken equal to the space-mean values P and T . The rate at which mass crosses the throat = Q , say, is in one-dimensional nozzle theory, neglecting energy losses,

$$Q = SP \left(\frac{\gamma}{RT} \right)^{\frac{1}{2}} \left(\frac{2}{\gamma+1} \right)^{[\gamma+1]/[2(\gamma-1)]}, \quad (10)$$

with an error involving ϵ , less than $6\frac{1}{2}\%$ for all ϵ up to 0.35. The most important ϵ terms in the formulae for nozzle flow have been given by Rateau (1932), or can be derived immediately from his results. Equation (10) can be written as

$$Q = \frac{\psi SP}{\sqrt{(RT)}}, \quad (11)$$

where

$$\psi = \gamma^{\frac{1}{2}} \left(\frac{2}{\gamma+1} \right)^{[\gamma+1]/[2(\gamma-1)]}. \quad (12)$$

This function ψ is insensitive to changes in γ , and it is found that ψ lies within 1 % of 0.659 for all Service propellants. In view of the omission of energy losses in the nozzle, it is preferable to leave ψ as an adjustable constant, which may vary with the nature of the leakage channels, but which should be the same for all charges in a given gun. ψ will be referred to as the 'nozzle discharge coefficient'. For recoilless guns with well-streamlined nozzles, ψ lies near 0.63. On the other hand, where the leakage passage is tortuous, as in the escape of gas past the shot in a worn gun, energy losses are greater and values of ψ of 0.6 or less seem to be suitable.

It is assumed that no unburnt cordite passes the nozzle throat. The equation for N becomes

$$C \frac{dN}{dt} = C \frac{d\phi}{dt} - \frac{\psi SP}{\sqrt{RT}}. \quad (13)$$

(3.4) Energy equation for the propellant gases

In time dt the gas gives kinetic energy $APdx$ to the shot, a mass $Cd\phi$ is evolved from the cordite, and $C(d\phi - dN)$ passes out through the nozzle. This process can be divided into three steps:

(i) The gas, of internal energy $NCE(T)$, provides kinetic energy $APdx$; in this stage the temperature changes by

$$dT = - \frac{APdx}{CNc_v}, \quad (14)$$

where c_v is the specific heat (at constant volume) of the gases, at temperature T ; since $c_v = R/(\gamma - 1)$, (14) can be written as

$$NdT = - \frac{(\gamma - 1)APdx}{CR}. \quad (15)$$

(ii) A mass $Cd\phi$, with internal energy $E(T_0)Cd\phi$, enters the gas; hence

$$Cd\phi E(T_0) = CNc_v dT + Cd\phi E(T),$$

and so

$$dT = \frac{d\phi}{N} (T_0 - T), \quad (16)$$

if c_v is constant over the range T to T_0 .

(iii) A mass $C(d\phi - dN)$ escapes through the nozzle. Since the expansion is adiabatic, then

$$d(\log T) = -(\gamma - 1) d \log \left(\frac{1}{\rho} - \eta \right)$$

or

$$\frac{dT}{T} = \frac{\gamma - 1}{[(1/\rho) - \eta]} \frac{d\rho}{\rho^2}.$$

Also

$$\frac{d\rho}{\rho} = \frac{dN - d\phi}{N}.$$

Hence

$$\frac{dT}{T} = \left(\frac{\gamma - 1}{1 - \eta\rho} \right) \left(\frac{dN - d\phi}{N} \right) = (\gamma - 1)(1 + \epsilon) \frac{(dN - d\phi)}{N}. \quad (17)$$

Summing all three effects (15), (16) and (17),

$$N \frac{dT}{dt} = -(\gamma-1) \frac{AP}{CR} \frac{dx}{dt} + (T_0 - T) \frac{d\phi}{dt} + (\gamma-1)(1+\epsilon) T \left(\frac{dN}{dt} - \frac{d\phi}{dt} \right),$$

and so
$$\frac{d}{dt}(NT) = -(\gamma-1) \frac{AP}{CR} \frac{dx}{dt} + T_0 \frac{d\phi}{dt} - \{\gamma + (\gamma-1)\epsilon\} T \left(\frac{d\phi}{dt} - \frac{dN}{dt} \right),$$

and, using (13),

$$\frac{d}{dt}(NT) = -(\gamma-1) \frac{AP}{CR} \frac{dx}{dt} + T_0 \frac{d\phi}{dt} - \{\gamma + (\gamma-1)\epsilon\} \frac{\psi SP \sqrt{(RT)}}{CR}. \quad (18)$$

Heat losses can be included in the way explained by Kent and Vinti. If the heat loss is approximated by $\chi \times \frac{1}{2} W V^2$, where χ is a constant, one can write

$$\bar{\gamma} - 1 = (1 + \chi)(\gamma - 1), \quad (19)$$

and should then replace $\gamma - 1$ by $\bar{\gamma} - 1$ in that term of (18) which involves dx/dt .

The term $(\gamma - 1)\epsilon$ can be omitted with an error which reaches 7% at $\epsilon = 0.35$, and is almost exactly compensated by the error in using a constant ψ . One may write, therefore,

$$\frac{d}{dt}(NT) = -(\bar{\gamma} - 1) \frac{AP}{CR} \frac{dx}{dt} + T_0 \frac{d\phi}{dt} - \frac{\gamma \psi SP \sqrt{(RT)}}{CR}. \quad (20)$$

(3.5) Summary of the equations

Four special points in the motion are distinguished: nozzle start, shot start, 'all burnt' and shot ejection. Quantities at these epochs are denoted by suffixes *NS*, *SS*, *B* and *E*. It is assumed that flow through the nozzle begins at a pressure to be known as the 'nozzle-start pressure'.

For nozzle open, shot in motion, and charge not completely burnt,

$$\left. \begin{aligned} P(U + Ax - C/\delta) &= CNRT, \\ W \frac{d^2x}{dt^2} &= AP, \\ D \frac{df}{dt} &= -\beta P^\alpha, \\ \phi &= (1-f)(1+\theta f), \\ \frac{dN}{dt} &= \frac{d\phi}{dt} - \frac{\psi SP}{C\sqrt{(RT)}}, \\ \frac{d}{dt}(NT) &= -(\bar{\gamma} - 1) \frac{AP}{CR} \frac{dx}{dt} + T_0 \frac{d\phi}{dt} - \frac{\gamma \psi SP \sqrt{(RT)}}{CR}. \end{aligned} \right\} \quad (21)$$

Various special cases are derived thus: before the nozzle opens, $S \equiv 0$; before shot start, $x \equiv 0$; after 'all burnt', $\phi \equiv 1$ and (5) is replaced by

$$P(U + Ax - CN\eta) = NCRT. \quad (22)$$

4. REDUCTION TO A NON-LEAKING GUN

In this section a simplified version of the set of equations (21) is dealt with, and the nature of the effects produced by gas leakage will appear by comparison with the analogous equations for an orthodox gun. It may be said at once that the approximations made are quite drastic, but that this has not prevented the working ballisticians, with his mass of empirical data, from founding on this basis a successful treatment of orthodox guns. A similar development for leaking guns appears to be possible.

It is assumed that the shot starts without initial resistance and that the nozzle flow is established at a low pressure. It is assumed, further, that the rate of burning is proportional to pressure, that is, that $\alpha = 1$, and that the leakage area S is constant.

Placing $S = 0$ in (21), one returns to the equations of an orthodox gun. The variation of the gas temperature with time is decided by the competition between the two terms remaining on the right of (20): the first term represents the lowering of the energy of the gas by the work done on the projectile; the second term corresponds to the increase of energy by the burning of the propellant. The competition between these terms leads, as is well known, to a gas temperature which falls only slowly from the initial value T_0 , though the decrease becomes more marked towards the end of the burning; after the cordite has been consumed there is of course a rapid drop of temperature during the adiabatic expansion of the gases.

When there is a constant leakage area S , there is an additional term on the right of (20), tending to lower the temperature by the work done in pushing gas out of the nozzle. The effect on the (temperature: time) relation depends chiefly on the ratio S/A , where A is the cross-sectional area of the bore. For recoilless guns with nozzles of good shape, S/A lies near 0.65, and it has been found that in such cases the temperature shows a rapid drop when the nozzle opens, thereafter flattens out, and later shows the increasing rate of decline characteristic of normal guns. For most of the period of burning, the gas temperature in the recoilless gun lies near 0.9 of the mean temperature in the corresponding period of an ordinary gun.

One can approximate to both leaking and sealed guns by equations in which the temperature is given a mean value during the period of burning. For a normal gun this mean should be about $0.9T_0$. The mean to be used with leakage depends chiefly on S/A , decreasing to about $0.8T_0$ at $S/A = 0.7$. Thus one effect of leakage may be expressed as a decrease in the effective force constant of the propellant, by as much as 10 % in a fully recoilless gun.

The mean value of RT is written as λ . Then (21) may be integrated to give

$$P(U + Ax - C/\delta) = \lambda C(1 - f)(1 - \Psi + \theta f), \quad (23)$$

where

$$\Psi = \frac{\psi SD}{\beta C \lambda^2}. \quad (24)$$

The dimensionless parameter Ψ is, as will be shown, the fundamental quantity expressing the effect of leakage on the internal ballistics. The values of Ψ found in

practice fall into two distinct classes. In the first, appropriate to smooth-bore guns and mortars, and leakage due to erosion, Ψ is of order 0.1. It will be seen that in this region the deviation of the ballistics from that of normal guns can be expressed as linear functions of Ψ . Larger values of Ψ do not appear in practice, at least not in the experience of the author, until values in the range 0.4–0.6 are reached. These are obtained in firings of recoilless guns with tubular propellant. When cord propellant is used, rather larger values of Ψ are found. However, cord is but little used in recoilless guns.

Equation (23) may be written as

$$P(U + Ax - C/\delta) = C(1 - \Psi)\lambda(1 - f)(1 + \theta'f), \quad (25)$$

where

$$\theta' = \theta/(1 - \Psi). \quad (26)$$

An orthodox gun with a charge $C(1 - \Psi)$ of propellant with mean force constant λ and form factor θ' , would have, instead of (25), the equation

$$P(U + Ax - C(1 - \Psi)/\delta) = C(1 - \Psi)\lambda(1 - f)(1 + \theta'f), \quad (27)$$

which differs from (25) in a term which is important only for high densities of loading. Indeed, the term $C(1 - \Psi)/\delta$ is correct only at the beginning of motion in an orthodox gun, this term later increasing to $C(1 - \Psi)\eta$, which is about 60 % greater. Thus, up to 'all-burnt' the leaking gun behaves almost as if it were an orthodox gun with *the same dimensions*, with the smaller charge $C(1 - \Psi)$, the numerically bigger form factor $\theta' = \theta/(1 - \Psi)$, and a force constant reduced as described in an earlier paragraph.

Since θ' is numerically greater than θ , it is clear that a propellant shape which is degressive in an orthodox gun is even more so in a leaking gun—hence the tendency not to use cord in recoilless guns. For the same reason, the virtues of a progressive burning shape are especially pronounced in recoilless guns.

It is also possible to approximate to the recoilless gun by a normal gun with the same charge C , form factor $\theta/(1 - \Psi)$, and smaller force constant $\lambda(1 - \Psi)$. However, the variation of force constant between different Service propellants is small, and it is therefore difficult to use one's previous ballistic experience to form an intuitive grasp of the effects of this substantial drop in λ .

$$\text{The 'effective charge' } C' = C(1 - \Psi) = C - \frac{\psi SD}{\beta \lambda^{\frac{1}{2}}}. \quad (28)$$

This may be recast into a form which brings out more clearly the significance of the separate terms. For

$$D(1 - f) = \frac{\beta W V}{A},$$

and so, writing V_B for the velocity when the cordite is just burnt,

$$\frac{D}{\beta} = \frac{W V_B}{A}.$$

For a typical gun, V_B is about 80–90 % of the muzzle velocity V_M , roughly

$$\frac{D}{\beta} = 0.85 \frac{WV_M}{A}. \quad (29)$$

For an orthodox gun with charge C' , the energy of the shot at the muzzle is

$$\frac{WV_M^2}{2} \approx \frac{C'RT_0}{3(\bar{\gamma}-1)} \approx \frac{4C'\lambda}{3}. \quad (30)$$

Substituting from (29) and (30) into (28),

$$C \approx \frac{3WV_M^2}{8\lambda} + 0.85 \frac{\psi SWV_M}{\lambda^{\frac{1}{2}}A}. \quad (31)$$

These terms may be regarded as the 'part that pushes' and the 'part that leaks', respectively. For a recoilless gun, $S/A \approx 0.65$, and $\psi \approx 0.63$, so that

$$C \approx \frac{3WV_M^2}{8\lambda} + 0.35 \frac{WV_M}{\lambda^{\frac{1}{2}}}. \quad (32)$$

The second term on the right may be regarded as the charge thrown backwards to balance the muzzle momentum of the shot, and implies that in the nozzles usually fitted to recoilless guns the specific impulse of the gases is about $3\lambda^{\frac{1}{2}}$.

A formula of type (32) was used by the engineers who designed the German recoilless guns, and was based on the argument that part of the charge pushes and the rest balances the recoil. Actually the coefficient of WV_M^2/λ was taken to be rather greater than $\frac{3}{8}$, and the coefficient of $WV_M/\lambda^{\frac{1}{2}}$ was determined by fitting to firings with 'reasonable' charges. It can be seen that this is not exact, though it does give a useful guide in preliminary design. The basic reduction to an orthodox gun holds only up to the end of burning. Afterwards, as will appear in § 5.3, the pressure-space curve shows a much more rapid drop than in an orthodox gun. It is clear that the factor $\frac{3}{8}$ in (32) ought to be increased, as was indeed done by the Germans. It is possible to develop further the line of approach represented by (32), in an empirical manner, and to obtain also corrections for the pressure gradient in the gun.

In an orthodox gun with a normal type of pressure-space curve, the muzzle velocity is, for small changes in charge, roughly proportional to the charge raised to the power n , where n depends on the details of the gun and charge but is usually around 0.7. At any rate, n is positive. Thus for small leakage

$$\text{muzzle velocity} \propto (1 - \Psi)^n \approx 1 - n\Psi.$$

The peak pressure may be treated similarly. For propellants with a nearly constant burning surface, and not too fast a rate of burning, the maximum pressure is proportional to the square of the charge; hence

$$\text{peak pressure} \propto (1 - \Psi)^2 \approx 1 - 2\Psi.$$

Exceptions to this rule are fast charges in which the peak pressure occurs at 'burnt', and charges of cord. These cases can also be treated by the reduction to a non-leaking gun.

The central ballistic parameter of orthodox ballistics is, omitting corrections for pressure gradient,

$$M = \frac{A^2 D^2}{\beta^2 C' W R T_0},$$

which for small Ψ , at constant charge, varies as

$$M \propto 1 + \Psi.$$

An important point is that these ballistic quantities are linear in Ψ , for small Ψ .

When considering not small Ψ but small changes in a large Ψ , it is found that such quantities as

$$\frac{1}{\text{muzzle velocity}} \frac{d(\text{muzzle velocity})}{d\Psi}$$

have a factor $1 - \Psi$ in their denominators. It follows that recoilless guns, where Ψ is of order 0.5, are more than normally sensitive to changes in web size or rate of burning.

5. METHODS OF SOLUTION

(5.1) Numerical integration

It is desirable to write equations (21) in terms of non-dimensional variables.

Write

$$\left. \begin{aligned} x &= x_0 \xi \quad \text{with} \quad x_0 = U/A, \\ S &= \mu A, \\ t &= \frac{x_0 \tau}{\sqrt{(RT_0)}}, \\ T &= T_0 T', \\ P(t) &= CRT_0 \Pi(\tau)/U, \\ V &= RT_0 \frac{d\xi}{d\tau}. \end{aligned} \right\} \quad (33)$$

Equations (21) become

$$\left. \begin{aligned} \Pi(1 + \xi - C/U\delta) &= NT', \\ \frac{d^2 \xi}{d\tau^2} &= \frac{C}{W} \Pi, \\ \frac{df}{d\tau} &= -\frac{\beta U}{DA \sqrt{(RT_0)}} \left(\frac{CRT_0}{U} \right)^\alpha \Pi^\alpha, \\ \phi &= (1-f)(1+\theta f), \\ \frac{dN}{d\tau} &= \frac{d\phi}{d\tau} - \mu \psi \Pi(T')^{-1}, \\ \frac{d}{d\tau}(NT') &= -(\bar{\gamma} - 1) \Pi \frac{d\xi}{d\tau} + \frac{d\phi}{d\tau} - \gamma \mu \psi \Pi(T')^{-1}. \end{aligned} \right\} \quad (34)$$

After 'all burnt',

$$\Pi \left(1 + \xi - \frac{C}{U} \eta N \right) = NT'. \quad (35)$$

Equations (34) can be solved by a step-by-step integration proceeding in steps of equal $\Delta\tau$. The method is not unusual in any way. It is possible to reduce the labour of computation by the following approximations: using Δ to denote increments over a step,

$$\Delta N = \Delta\phi - \psi\mu(T')^{-1} \Delta \left\{ \int \Pi d\tau \right\} \quad (36)$$

and
$$\Delta(NT') = -\frac{(\bar{\gamma}-1)W}{2C} \Delta \left\{ \left(\frac{d\xi}{d\tau} \right)^2 \right\} + \Delta\phi - \gamma\psi\mu(T')^{-1} \Delta \left\{ \int \Pi d\tau \right\}. \quad (37)$$

Bars denote averages over the step. Since $(T')^{\frac{1}{2}}$ varies only slowly, it is possible to use the same average values for several steps in succession; the most convenient number of such steps has been found to be two, usually, but sometimes blocks of four are more convenient.

The value of 'reduced variables' in this connexion is that they reduce the differences between various guns. By thinking in these terms the underlying resemblances become more obvious.

Numerical integration can be used when the nozzle throat area S varies during the firing; the most important application where this occurs is to the worn orthodox gun. A considerable number of solutions have been carried out for various forms of bore resistance and leakage area depending on travel in various ways, and the author's experience has shown that a good approximation to the effect of gas leakage in this case can be obtained by using the method of § 4 with a mean area S equal to half the maximum leakage area between the unengraved projectile and the worn gun.

(5.2) Solution with linear rate of burning

If the rate of burning of the propellant is proportional to the pressure, it is possible to find an approximate analytical solution for the period up to 'all burnt'. There is a considerable saving of time over the previous method of numerical integration.

Equations (21) are used with the additional assumptions: the rate of burning is proportional to pressure, i.e. $\alpha = 1$; the nozzle throat area S is independent of time. In this method it is convenient to use ordinary units except that the temperature T is best replaced by $T_0 T'$ as in the previous method. The equations are

$$D \frac{df}{dt} = -\beta P, \quad (38)$$

$$\phi = (1-f)(1+\theta f), \quad (39)$$

$$P(U - C/\delta + Ax) = CNRT_0 T', \quad (40)$$

$$WV \frac{dV}{dx} = AP, \quad (41)$$

$$\frac{dN}{dt} = \frac{d\phi}{dt} - \frac{\psi SP}{C\sqrt{(RT_0)}} (T')^{-1}. \quad (42)$$

Let suffixes *NS* and *SS* denote conditions at nozzle start and shot start respectively.

Eliminating P from (38) and (42) and integrating,

$$N = \phi + \frac{\psi SD}{\beta C \sqrt{(RT_0)}} \frac{(f - f_{NS})}{\nu_1}, \quad (43)$$

where

$$\frac{1}{\nu_1} = \frac{1}{f - f_{NS}} \int_{f_{NS}}^f (T')^{-1} df.$$

It is convenient, as in § 4, to write

$$\Psi = \frac{\psi SD}{\beta C \sqrt{(RT_0)}}, \quad (44)$$

so that (43) becomes

$$N = \phi - \frac{\Psi(f_{NS} - f)}{\nu_1}. \quad (45)$$

Inspection of a number of accurate solutions has shown that it is sufficiently correct to take

$$\nu_1 = \frac{1}{4}(1 + 3\sqrt{T'}). \quad (46)$$

From (38) and (41),

$$V = \frac{AD}{\beta W} (f_{SS} - f). \quad (47)$$

This holds from shot start to 'all burnt'.

From (20), (38) and (41),

$$NRT = (NRT)_{NS} + RT_0(\phi - \phi_{NS}) - (\bar{\gamma} - 1) \frac{WV^2}{2C} + \gamma \Psi RT_0(f - f_{NS}) \nu_2, \quad (48)$$

where

$$\nu_2 = \frac{1}{f - f_{NS}} \int_{f_{NS}}^f (T')^{\frac{1}{2}} df.$$

Since $N_{NS} = \phi_{NS}$ and $(RT)_{NS} = RT_0$ (if $P_{NS} \leq P_{SS}$) $\simeq RT_0$ (if $\dot{P}_{NS} > P_{SS}$), it follows that

$$NT' = \phi - \frac{(\bar{\gamma} - 1) WV^2}{2CRT_0} + \gamma \Psi(f - f_{NS}) \nu_2,$$

which with (45) reduces to

$$T' \left\{ \phi + \frac{(\gamma - 1) \Psi(f_{NS} - f)}{\nu_3} \right\} = \phi - \frac{(\bar{\gamma} - 1) WV^2}{2CRT_0}, \quad (49)$$

where

$$\frac{\gamma - 1}{\nu_3} = \frac{\gamma \nu_2}{T'} - \frac{1}{\nu_1}.$$

Equation (49) holds after nozzle opening. If the shot starts before this, (49) is used, until the nozzle opens, with $\Psi = 0$.

A relation is also needed between the travel and the fraction of charge burnt. Eliminating P from (40) and (41),

$$\frac{A dx}{U - C/\delta + Ax} = \frac{WV dV}{CNRT_0 T'} = - \left(\frac{AD}{\beta W} \right)^2 \frac{W}{CRT_0} \frac{(f_{SS} - f) df}{NT'},$$

so that

$$\int \frac{A dx}{U - C/\delta + Ax} = - \left(\frac{AD}{\beta} \right)^2 \frac{1}{WCRT_0 \nu_4} \int \frac{(f_{SS} - f) df}{N}, \quad (50)$$

where ν_4 is a certain average reduced temperature.

The dimensionless quantity $A^2 D^2 / \beta^2 W C R T_0$ corresponds to the central ballistic parameter of the classical treatises on internal ballistics.

Equation (50) is true whether shot start occurs before or after the opening of the nozzle. One must now, however, distinguish between these two cases.

Case A. $P_{NS} \leq P_{SS}$

In this case N is given by (45) for the whole of the period of motion of the shot. Substituting into (50) and integrating from shot start onwards,

$$\ln \left\{ 1 + \frac{Ax}{U - C/\delta} \right\} = \left(\frac{AD}{\beta} \right)^2 \frac{I}{W C R T_0 \nu_4}, \quad (51)$$

where

$$I = \int_f^{f_{ss}} \frac{(f_{ss} - f) df}{\phi + \Psi(f - f_{NS})/\nu_1}.$$

In the integrand ν_1 is a function of T' , and the integration can be carried out only if ν_1 is averaged. It is found that, writing the average value of Ψ/ν_1 as Ω :

For $\theta \neq 0$:

$$I = \frac{1}{2\theta} \ln \left\{ \frac{1 - \Omega f_{NS} + (\theta + \Omega - 1)f_{SS} - \theta f_{SS}^2}{1 - \Omega f_{NS} + (\theta + \Omega - 1)f - \theta f^2} \right\} \\ + \frac{\theta + \Omega - 1 - 2\theta f_{SS}}{\theta [4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \left[\tanh^{-1} \left\{ \frac{2\theta f - (\theta + \Omega - 1)}{[4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \right\} \right. \\ \left. - \tanh^{-1} \left\{ \frac{2\theta f_{SS} - (\theta + \Omega - 1)}{[4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \right\} \right]. \quad (52)$$

This holds if the arguments of the \tanh^{-1} terms are less than unity. Where any argument is greater than unity, \tanh^{-1} should be replaced by \coth^{-1} .

For $\theta = 0$:

$$I = \frac{f_{SS} - f}{1 - \Omega} - \frac{1 - \Omega f_{NS} - (1 - \Omega)f_{SS}}{(1 - \Omega)^2} \ln \left\{ \frac{1 - \Omega f_{NS} - (1 - \Omega)f}{1 - \Omega f_{NS} - (1 - \Omega)f_{SS}} \right\}. \quad (53)$$

It has been found that pressures correct to a few per cent can be obtained by taking

$$1 - \nu_4 = \frac{50(1 - T')^2(T' - 0.78)}{1 + 2.5(1 - T')}, \quad (54)$$

$$\Omega = \frac{2\Psi}{3\sqrt{T'} - 1}. \quad (55)$$

The fact that ν_4 is greater than unity for T' less than 0.78 is due to Ω being least accurate in this region. The pair of formulae, (54) and (55), have been chosen to compensate for each other's deficiencies even at these low temperatures. The integral I becomes infinite for $\Omega = 1$, and changes in Ω produce very large changes in I if Ω is near 1. This approximate method is unreliable for Ω greater than 0.8.

Case B. $P_{NS} < P_{SS}$

Before the opening of the nozzle $N = \phi$, and so

$$\ln \left\{ 1 + \frac{Ax}{U - C/\delta} \right\} = \left(\frac{AD}{\beta} \right)^2 \frac{1}{WCRT_0 \nu_4} I(\Omega = 0). \quad (56)$$

After nozzle opening N is taken from (45) and

$$\ln \left\{ \frac{U - C/\delta + Ax}{U - C/\delta + Ax_{NS}} \right\} = \left(\frac{AD}{\beta} \right)^2 \frac{J}{WCRT_0 \nu_4}, \quad (57)$$

where

$$J = \int_f^{f_{NS}} \frac{(f_{SS} - f) df}{\phi + \Psi(f - f_{NS})/\nu_1}.$$

This integral can be evaluated as:

For $\theta \neq 0$:

$$J = \frac{1}{2\theta} \ln \left\{ \frac{1 + (\theta - 1)f_{NS} - \theta f_{NS}^2}{1 - \Omega f_{NS} + (\theta + \Omega - 1)f - \theta f^2} \right\} \\ + \frac{\theta + \Omega - 1 - 2\theta f_{SS}}{\theta [4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \left[\tanh^{-1} \left\{ \frac{2\theta f - (\theta + \Omega - 1)}{[4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \right\} \right. \\ \left. - \tanh^{-1} \left\{ \frac{2\theta f_{NS} - (\theta + \Omega - 1)}{[4\theta(1 - \Omega f_{NS}) + (\theta + \Omega - 1)^2]^{\frac{1}{2}}} \right\} \right]. \quad (58)$$

For $\theta = 0$:

$$J = \frac{f_{NS} - f}{1 - \Omega} - \frac{1 - \Omega f_{NS} - (1 - \Omega)f_{SS}}{(1 - \Omega)^2} \ln \left\{ \frac{1 - \Omega f_{NS} - (1 - \Omega)f}{1 - f_{NS}} \right\}, \quad (59)$$

with the usual caution about replacement of \tanh^{-1} by \coth^{-1} if the argument becomes greater than unity.

A summary of the use of these equations is now given. Various mean values such as Ω and the ν 's have been introduced, and rules for their estimation will be given. These rules were derived by analysis of a number of accurate runs for various conditions, and it is not suggested that they are the only rules which will work. However, these rules are simple, and experience has shown that pressures, velocities and travels can be computed in this way to within 2 or 3 %. This is ample accuracy for routine ballistic calculations, being usually less than the error arising from uncertainties in the rate of burning and in the nozzle-start and shot-start pressures. The computation, which is of self-consistent nature, converges very rapidly, unless Ω approaches 0.8. For this case an alternative method has been devised by Mr T. Vickers; this method uses the fact that the pressure varies only slowly with the travel.

Case A. $P_{NS} \leq P_{SS}$

Solution at nozzle start. Equation (40) with $P = P_{NS}$, $T' = 1$, and $x = 0$ gives $N_{NS} = \phi_{NS}$. f_{NS} follows from (39).

Solution at shot start. Guess T'_{SS} . (40), with $P = P_{SS}$ and $x = 0$, gives N_{SS} . Calculating ν_1 from (46), ϕ_{SS} and f_{SS} follow from (39) and (45). Take, in this part of the computation,

$$\nu_3 = 1 - \frac{6(1 - T')}{2 + 1000(1 - T')^3}, \quad (60)$$

and obtain T'_{SS} from (49) with $V = 0$. The process is now repeated until a self-consistent set of values is reached.

Solution for any value of f smaller than f_{SS} . Guess T' at this value of f . In succession ν_1 is obtained from (46), ϕ from (39), N from (45), and V from (47). Estimate ν_3 here as

$$\nu_3 = 1.75T' - 0.75, \quad (61)$$

and then T' can be calculated from (49). This is repeated until the results are self-consistent.

To calculate the corresponding travel use (54), (55) and (51), with either (52) or (53). The pressure can now be calculated from (40).

Maximum pressure. By calculating the pressure at three or four evenly spaced values of f , the maximum pressure can be found by interpolation.

A discontinuity of pressure may appear at 'all burnt' owing to the transition from

$$P(U - C/\delta + Ax) = CNRT$$

to

$$P(U - \eta CN + Ax) = CNRT.$$

The discontinuity is usually small and can be smoothed out. A large discontinuity indicates that the approximation involved in the replacement of (4) by (5) in § 3.2 is not sufficiently accurate in this particular example, and that a corrected value of RT_0 must be used up to 'all burnt', as explained in § 3.2.

Case B. $P_{NS} > P_{SS}$

Solution at shot start. Equation (40) with $P = P_{SS}$, $x = 0$ and $T' = 1$, gives $N_{SS} = \phi_{SS}$, and f_{SS} follows from (39).

Solution at nozzle start. Guess f_{NS} , obtaining V from (47), $\phi = N$ from (39), T' from (53) with $\Psi = 0$, and x from (56) with ν_4 from (54).

The pressure can be calculated from (40). Repeat with other values of f and interpolate for $P = P_{NS}$.

Solution after nozzle start. The only difference from the procedure in case A is that (58) and (59) replace (52) and (53).

(5.3) After 'all burnt'

This case is so much simpler that an approximate analytical solution can be derived by modifying the method used for the same period in the orthodox gun.

From the exact equations, in reduced units,

$$\frac{d}{d\tau}(NT') = -\frac{(\bar{\gamma} - 1)NT'}{(1 + \xi - \eta CN/U)} \frac{d\xi}{d\tau} - \frac{\gamma\psi\mu(T')^{\frac{1}{2}}NT'}{(1 + \xi - \eta CN/U)},$$

so that

$$\frac{d(NT')}{NT'} = - \frac{K d\xi}{1 + \xi - \eta CN/U}, \quad (62)$$

where

$$K = \bar{\gamma} - 1 + \frac{\gamma \mu \psi (T')^{\frac{1}{2}}}{d\xi/d\tau}. \quad (63)$$

Let suffix B denote conditions at 'all burnt', $\phi = 1$, and E those at shot ejection. Integrate (62) by taking mean values of K and $\eta CN/U$, giving

$$\log \left\{ \frac{NT'}{(NT')_B} \right\} = -\bar{K} \log \left\{ \frac{1 + \xi - \eta CN/U}{1 + \xi_B - \eta CN/U} \right\}. \quad (64)$$

One can write (62) in the alternative form

$$\frac{d}{d\tau}(NT') = -K \frac{d\xi}{d\tau} \Pi = -K \frac{W}{C} \frac{d\xi}{d\tau} \frac{d^2\xi}{d\tau^2},$$

so that

$$\left(\frac{d\xi}{d\tau} \right)^2 - \left(\frac{d\xi}{d\tau} \right)_B^2 = \frac{2C}{\bar{K}W} \{ (NT')_B - NT' \}. \quad (65)$$

Finally,

$$N_B - N = \mu \psi \int_B (T')^{-1} \Pi d\tau = \mu \psi (T')^{-1} \left\{ \frac{d\xi}{d\tau} - \left(\frac{d\xi}{d\tau} \right)_B \right\}. \quad (66)$$

From these formulae approximate values of pressure and velocity can be calculated for any position of the shot after 'all burnt'. In general one needs only the muzzle velocity, for which the detailed working runs thus. ξ_B , T'_B , $\left(\frac{d\xi}{d\tau} \right)_B$ and N_B are known, from which K_B is calculated. Guess T' , $d\xi/d\tau$ and N at the muzzle, whose value of ξ is known. K_E follows from (63), and then one may take

$$\bar{K} = \frac{1}{2}(K_B + K_E), \quad \bar{N} = \frac{1}{2}(N_B + N_E),$$

thence obtaining $(NT')_E$ from (64). Substituting in (65), one finds $(d\xi/d\tau)_E$. Finally, N_E is obtained from (66), using the mean value

$$(T')^{-1} = \frac{1}{2}\{(T'_B)^{-1} + (T'_E)^{-1}\}.$$

The cycle is now repeated until self-consistent. The number of cycles necessary depends on the experience of the computer, but is normally of order 3.

Errors in $(d\xi/d\tau)_E$, $(N)_E$ and $(T')_E$ in a typical calculation of this type were found to be 0.2, -4 and -16 % respectively of the change from 'burnt' to 'ejection', i.e. about 0.1, -2 and -5 % respectively of the value at ejection. The error in muzzle velocity is only a few ft./sec., which is a trivial price to pay for the reduction of computation effected by this method.

It can be seen that the shape of the pressure-space curve during the adiabatic expansion, which in ordinary guns is determined by $\bar{\gamma} - 1$, is settled here by the quantity \bar{K} . For a typical recoilless gun, \bar{K} is of order 2, whereas $\bar{\gamma} - 1$ is only about 0.3; hence after 'all burnt' the pressure-space curve falls much more rapidly than in an orthodox gun.

I am indebted to Colonel G. H. Hinds and Professor N. F. Mott for their encouragement of this investigation, to the Chief Scientific Officer, Ministry of Supply, for permission to publish this paper, to Dr H. H. M. Pike and Mr S. W. Coppock for helpful discussions on internal ballistics, and to Mrs H. N. Wilkinson and Mr T. Vickers for their assistance in the application of these methods.

REFERENCES

- Cooke, C. J. 1921 U.S. Patent 1,380,358.
Cranz, C. 1926 *Ballistik*, 2. Berlin.
Rateau, A. 1932 *Mém. Artill. fr.* 11, 5.
-

The internal ballistics of a gun after shot ejection

By J. CORNER, *Armament Research Department, Ministry of Supply*

(Communicated by N. F. Mott, F.R.S.—Received 4 April 1946)

The motion of gas inside a gun barrel after the ejection of the shot is studied theoretically, and it is shown that a fairly simple solution is possible for muzzle velocities of order 3000 ft./sec. Formulae are obtained for the pressure at the breech and the rate of outflow, for times after shot ejection, and comparison with an experimental pressure-time record shows good agreement. It is shown that by simple modification the classical work of Hugoniot and Rateau can be made sufficiently accurate for many practical problems. Some of the applications of this work are indicated.

1. INTRODUCTION

The problem of the motion of the gas in a gun after the shot has left is rather similar to Lagrange's ballistic problem (cf. Love & Pidduck 1921), and is connected with the latter by the fact that the final state in Lagrange's problem, with the shot just at the muzzle, is the initial state in the present work. The emptying of a gun was apparently first discussed by Hugoniot (1886), in a form which has been repeated in many other applications to the emptying of a reservoir of gas. As regards their use in internal ballistics, Hugoniot's results show how the breech pressure dies away after shot ejection and hence, by integration, how the gas increases the momentum of the recoiling parts after the shot has left the gun. However, for a long time these results were not applied in the design of recoil systems, which was carried out with the help of empirical factors drawn from accumulated experience. The use of a baffle outside the muzzle of a gun, to provide a forward thrust after the shot has emerged, was first proposed in the 1880's. Such a 'muzzle brake' is shown in figure 1, which is based on a modern practical design. The matter was first taken up seriously by the French in the 1914-18 war. It was found that it was necessary to pay more

attention to the theoretical details of the outflow from the gun, partly because of the larger number of empirical factors needed to cover the more complicated behaviour, and also because of the lack of empirical knowledge of such systems. With this aim, Rateau (1919, 1932) improved Hugoniot's work by (i) including the effect of the covolume, and (ii) correcting for the initial distribution of gas velocities along the barrel at the moment of shot ejection. Rateau's theory will be sketched later.

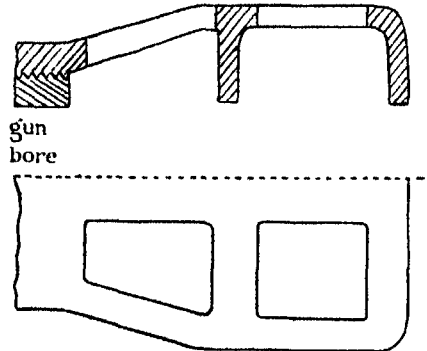


FIGURE 1. Section and external view of a typical muzzle brake with two baffles.

During the late war the use of muzzle brakes became almost universal on certain classes of guns, and theoretical calculations were made on various aspects of their behaviour. The work reported in this paper formed part of the programme, and arose from a desire to check the Rateau-Hugoniot formulae in a typical application. It was hoped that one might be able to explain certain apparently anomalous results which were being obtained in the experimental development. It was found, however, that Rateau's formulae are sufficiently accurate for the design of most guns which would use muzzle brakes. The results of the present paper have been applied to several other problems of gun ballistics.

The results given here do not cover the case of a low muzzle velocity, where the hydrodynamic behaviour is relatively more complicated. This is a region where the forces acting after ejection are smaller compared with the total momenta concerned, so that our results do cover the range of practical interest.

Terms have been neglected of order $(\text{charge weight}/\text{shot weight})^2$, which is about 0.2 for a typical anti-tank gun. The results are therefore not necessarily true for hypervelocity guns with roughly equal weights of charge and projectile. It should be noted that the initial state in the problem is itself not known to terms of this order, except in the limiting case where the propellant burns very rapidly.

2. NOTATION

It will be assumed that the internal volume of the gun has the form of a circular cylinder of cross-sectional area S and volume U . In applications to real guns U will be the total capacity of the gun, and S the bore area. The length of the real gun

will be somewhat less than U/S , because of the larger cross-section of the chamber, but the error in the results will be small for all normal guns. The size of the error will be discussed later.

Let the charge mass be C , shot mass W , and muzzle velocity V_0 .

Let x denote distance from the breech, t the time from the instant of shot ejection. At this point and time, the pressure is p , velocity of gas v , temperature T , and density ρ . It is convenient to introduce non-dimensional units of distance and time, defined by

$$\xi = \frac{xS}{U}, \quad \tau = \frac{V_0 St}{U}. \quad (1)$$

In terms of ξ and τ the results take a simpler form.

3. INITIAL STATE OF THE GAS IN THE BARREL

Assume that the charge is completely burnt before the shot reaches the muzzle. At $t = 0$ the shot is at the muzzle, and the velocity and pressure distribution along the barrel is the 'final state' of Lagrange's ballistic problem. It has been shown by Hicks & Thornhill (unpublished work) that the 'Lagrange approximation' can be modified in its assumptions so as to give the usual results to terms of order C/W , and a close approximation to Pidduck's limiting solution (Love & Pidduck 1921) when keeping terms of order $(C/W)^2$. The assumptions of Hicks & Thornhill are especially convenient for our purpose. For $C/W = \frac{1}{2}$, corresponding to muzzle velocities of order 3000 ft./sec., the $(C/W)^2$ term in the ratio of breech pressure to pressure at shot, is thought to be of order 3 % of the C/W term; therefore terms of order $(C/W)^2$ will be neglected throughout these calculations. It is, of course, true that the $(C/W)^2$ terms, certainly small in the initial state, may increase relatively to the other terms as time goes on: this seems unlikely, since the terms of order C/W do not increase very much, and later they die down. Covolume will not be neglected, which would indeed not be very plausible in the practical example discussed later, for which the breech pressure at shot ejection is about 10 tons/sq.in. Let the covolume be denoted by η . It is useful to introduce a ratio

$$\alpha = \frac{\eta}{(U/C) - \eta}, \quad (2)$$

which is related to the pressure at the instant of shot ejection. There are practical limits on the muzzle pressure, and α is therefore rarely more than 0.2. Certain terms will be expanded as power series in α , neglecting terms of order α^2 . The error will be of the same order as the neglect of $(C/W)^2$ terms. The terms of order α continually decrease, as would be expected.

Let T_0 be the gas temperature at the instant of shot ejection, averaged over the whole of the gas. Let R be the gas constant per gram of propellant gas, equal to the gas constant per g.mol. divided by the average molecular weight of the gas. R therefore varies slightly with temperature, owing to the variation in the minor

products of explosion; the variation in the amounts of the chief products CO , N_2 , CO_2 , H_2 and H_2O comes mainly from change in the water-gas equilibrium, which does not affect the average molecular weight. To the accuracy of present calculations R can be taken as a constant.

Let γ be the ratio of specific heats for the propellant gases. It depends on temperature, and a mean value must be taken, appropriate to the temperature range being considered.

It is assumed that the covolume is independent of temperature. This is true to the accuracy needed in these calculations. It follows then from the equation of state of the propellant gases, namely,

$$p\left(\frac{1}{\rho} - \eta\right) = RT, \quad (3)$$

that the equation of the adiabatics is

$$p\left(\frac{1}{\rho} - \eta\right)^\gamma = \text{constant}. \quad (4)$$

Hicks & Thornhill assume that $p(1/\rho - \eta)^\gamma$ is independent of position along the barrel, though of course it is a function of time. Their results for the instant of shot ejection are

$$p = \frac{RT_e}{(U/C) - \eta} \left[1 + \frac{C}{6W} (1 - 3\xi^2) \right], \quad (5)$$

$$\frac{1}{\rho} - \eta = \left(\frac{U}{C} - \eta \right) \left[1 + \frac{C}{6\gamma W} (3\xi^2 - 1) \right], \quad (6)$$

$$T = T_e \left[1 + \frac{(\gamma - 1)C}{6\gamma W} (1 - 3\xi^2) \right], \quad (7)$$

$$p\left(\frac{1}{\rho} - \eta\right)^\gamma = RT_e \left(\frac{U}{C} - \eta \right)^{\gamma-1}, \quad (8)$$

$$\text{and for the velocity distribution} \quad v = V_0 \xi. \quad (9)$$

These equations give the pressure, density and velocity distributions in terms of the muzzle velocity, gun dimensions, covolume and RT_e . The latter can be calculated from the energy of the propellant explosion, the energy of the shell as it leaves the muzzle, and the heat loss to the barrel; that is to say, in terms of muzzle velocity and known properties of the propellant. In particular, these results and those which follow can be applied without needing a solution of the ordinary equations of internal ballistics.

The equations (5)–(9) are true for the instant $t = 0$ only and are correct only to terms of order C/W .

The frictional pressure drop down the gun is neglected in these formulae and throughout the present work.

4. EARLIER WORK

The problem of the emptying of a reservoir of perfect gas by expansion through a nozzle was first treated by Hugoniot (1886), by assuming that the state of flow at any instant was the same as would be set up in steady flow with the reservoir pressure existing at that instant. This hypothesis of quasi-steady flow is plausible provided the reservoir pressure is not falling too rapidly. Rateau (1919, 1932) used the same assumption about quasi-steady flow, and introduced the covolume into the calculations. He assumed that the gun barrel acted as a convergent nozzle with throat area that of the bore, so that the gas velocity at the muzzle was equal to the local velocity of sound. He pointed out that the gun was an unusual type of reservoir, with its cross-section no bigger than the throat area of the nozzle; also the initial state, at the instant that flow through the nozzle began, was a non-uniform distribution of pressure and velocity down the reservoir. He suggested that this initial state could be corrected for in the following way. The gas was supposed to start from rest at an effective pressure p_i and temperature T_i , such that by adiabatic expansion the gas arrived at the true pressure p_e , right temperature T_e , and correct kinetic energy for the instant when the shot passes the muzzle. This statement is not unambiguous, with regard to such questions as the relation of breech and mean pressures, for example. The theory will be compared with experiment in § 8. The effective pressure p_i is obviously too high at the instant of ejection, with the idea of improving the form of the pressure/time curve as a whole. Unfortunately, the theoretical pressure stays too high all along. This error can be eliminated by not using Rateau's effective pressure and temperature, using instead the true values for the instant of ejection. The Rateau formulae are then simpler and easier to use than those obtained in the present paper, and while not quite so accurate are sufficient for many practical purposes. For this reason we summarize Rateau's results here.

The kinetic energy of the gases as the shot leaves is $\frac{1}{2}CV_0^2$. The 'effective temperature' T_i is therefore given by

$$RT_i \cdot RT_e = \frac{(\gamma - 1)}{6} V_0^2. \quad (10)$$

In this equation γ is a mean for the temperature range T_e to T_i . The difference between these temperatures is of order 100° C, and therefore γ can be taken as the value for T_e .

Rateau takes the covolume η to be 0.95 c.c./g. The effective density ρ_i follows from the true mean density C/U and the adiabatic equation, and is given by

$$\frac{1}{\rho_i} - 0.95 = \left(\frac{U}{C} - 0.95 \right) \left(\frac{T_e}{T_i} \right)^{1/(\gamma-1)}, \quad (11)$$

where ρ_i and C/U are in g./c.c. A time θ is introduced, defined by

$$\theta = \frac{2}{\gamma - 1} \frac{U}{S} \left[\frac{1}{\gamma} \left(\frac{\gamma + 1}{2} \right)^{(\gamma+1)/(\gamma-1)} \frac{[(1/\rho_i) + 1]^{\gamma-1}}{p_e [(U/C) - 0.95]^\gamma} \right]^{\frac{1}{2}}. \quad (12)$$

This time is characteristic of the time scale of the phenomena. The pressure at any time t measured from ejection is

$$p = p_e \left(\frac{U}{C} - 0.95 \right)^\gamma \left[\left(\frac{1}{\rho_i} + 1 \right) \left(1 + \frac{t}{\theta} \right)^{2/(\gamma-1)} - 1.95 \right]^{-\gamma}. \quad (13)$$

To get best agreement with experiment, one can take the 'pressure p_e ' to be the mean pressure in the barrel at the instant of ejection, and the 'pressure p ' of (13) to be the breech pressure. It is better still to omit the correction, leading to the following 'modified Rateau' formulae:

$$\theta = \frac{2}{\gamma-1} \frac{U}{S} \left[\frac{1}{\gamma} \left(\frac{\gamma+1}{2} \right)^{(\gamma+1)/(\gamma-1)} \left(\frac{(U/C)+1}{(U/C)-0.95} \right)^{\gamma-1} \frac{1}{RT_e [1 + (C/6W)]} \right]^\dagger, \quad (14)$$

and breech-pressure p at time t :

$$p = RT_e \left(\frac{U}{C} - 0.95 \right)^{\gamma-1} \left(1 + \frac{C}{6W} \right) \left[\left(\frac{U}{C} + 1 \right) \left(1 + \frac{t}{\theta} \right)^{2/(\gamma-1)} - 1.95 \right]^{-\gamma}. \quad (15)$$

Omitting the covolume terms, one obtains Hugoniot's results:

$$\theta = \frac{2}{\gamma-1} \frac{U}{S} \left[\frac{1}{\gamma} \left(\frac{\gamma+1}{2} \right)^{(\gamma+1)/(\gamma-1)} \frac{1}{RT_e [1 + (C/6W)]} \right]^\dagger \quad (16)$$

and

$$p = \frac{CRT_e}{U} \left(1 + \frac{C}{6W} \right) \left(1 + \frac{t}{\theta} \right)^{(-2\gamma)/(\gamma-1)}. \quad (17)$$

This formula for θ may be encountered in the literature with a rather different C/W term. This comes from the ambiguity in the choice of 'reservoir pressure', which in these simple theories is sometimes taken as breech pressure, sometimes as mean pressure.

5. DISCUSSION OF THE PROBLEM

Assume that each element of gas expands adiabatically, so that $p[(1/\rho) - \eta]^\gamma$ is independent of time and depends only on the element considered. Now on the assumptions of Hicks & Thornhill, this quantity is the same for all elements at the moment of shot ejection, and so

$$p \left(\frac{1}{\rho} - \eta \right)^\gamma = RT_e \left(\frac{U}{C} - \eta \right)^{\gamma-1}, \quad (18)$$

for all values of x and t . The assumption of adiabatic expansion after the instant of shot ejection involves two factors: (i) heat loss to the barrel is neglected after shot ejection; (ii) it is assumed that no shock waves are set up in the gas inside the gun. Of these, the second is certainly justifiable. As the shot passes out of the muzzle a rarefaction wave passes into the barrel and down towards the breech, and this wave cannot give rise to a discontinuity of density:

The heat loss to the barrel after ejection is not so easily dismissed, but it is difficult to include it in the equations without intolerable complications. The magnitude of

this heat loss may be compared with the heat loss before shot ejection. The velocities and times involved are of the same order in the two cases, while the densities after shot ejection are lower than before, thus reducing the heat loss. It is clear that the heat loss to the barrel after ejection will be smaller than the heat transfer while the shot is in the gun, though the orders of magnitude will be the same. If the effects of heat loss were to be included, the agreement with experiment would in fact be less good. This will be shown in § 8, and is due to partial compensation with some other approximations in our treatment.

Since $p[(1/\rho) - \eta]^\gamma$ is the same for all elements at any given time, the Eulerian equations of hydrodynamics may be used. The equation of continuity is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0, \quad (19)$$

and the equation of motion is
$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = -\frac{1}{\rho} \frac{\partial p}{\partial x}. \quad (20)$$

The velocity of sound at position x and time t is

$$c = \sqrt{\frac{dp}{d\rho}} = \frac{(\gamma RT_c)^{\frac{1}{2}}}{1 - \eta \rho} \left(\frac{(U/C) - \eta}{(1/\rho) - \eta} \right)^{\frac{1}{2}(\gamma-1)}. \quad (21)$$

The influence of the conditions outside the muzzle section is transmitted through the gas towards the breech with velocity $c - v$ relative to the barrel. This is assured by the fact that it is a rarefaction wave which is passing through the gas, not a shock wave. Let Z be the value of $\xi (= xS/U)$ at the point reached by the front of the rarefaction wave at time t ; that is, for $\xi < Z$ the behaviour of the gas is settled by the state of the gas inside the gun at the instant of shot ejection, while for $\xi > Z$ it is necessary to take into account conditions outside the muzzle. The motion of the wave front satisfies the equation

$$\frac{U}{S} \frac{dZ}{dt} = v - c, \quad (22)$$

the right-hand side being evaluated for $\xi = Z$.

It will be clear that the motion of the gas is simplest in the region $\xi < Z$. A solution for the region $\xi < Z$ will first be obtained, and used to integrate (22), thereby showing how far the rarefaction wave from the muzzle has progressed at any time. In § 9 the nature of the flow in the zone $\xi > Z$ will be discussed. The solution for $\xi < Z$ is, of course, not applicable at the muzzle except for guns of very high velocity, so it is not of much use by itself for the calculation of rates of outflow; for these the wave of rarefaction is an essential feature. However, this wave takes a relatively long time to reach the breech; for example, the breech pressure of the example given later falls to about a third of its value at ejection before the rarefaction wave hits the breech. Thus questions of recoil can often be settled by using the simpler solution up to this time and some rough approximation for the breech pressure thereafter. The extent to which this is applicable to any particular gun will be seen from later formulae.

6. SOLUTION BEFORE ARRIVAL OF WAVE FROM MUZZLE

Assume that the gas velocity v is a linear function of distance along the gun, and the pressure a quadratic function. These forms are suggested by the initial distributions at the instant of shot ejection. To attempt a solution on these lines write

$$v = V_0 \xi f(\tau) \left\{ 1 + \alpha g(\tau) + \frac{C}{W} h(\tau) \right\},$$

and
$$\frac{1}{\rho} = \left(\frac{U}{C} - \eta \right) k(\tau) \left[1 + \alpha l(\tau) + \frac{C}{W} m(\tau) + \xi^2 \left\{ n(\tau) + \alpha p(\tau) + \frac{C}{W} q(\tau) \right\} \right],$$

where the functions $f(\tau), \dots, q(\tau)$ are to be chosen to satisfy (18), (19) and (20) for all τ , and the initial conditions (5)–(9) at $\tau = 0$. It is possible to satisfy all these equations with errors of order α^2 , $\alpha C/W$ and $(C/W)^2$. The process is simple, although the details are a little laborious, and the solution is

$$v = \frac{V_0 \xi}{1 + \tau} \left[1 + \frac{C}{W} \frac{RT_e}{(2 - \gamma) V_0^2} \{ (1 + \tau)^{1 - \gamma} - (1 + \tau)^{-1} \} \right], \quad (23)$$

$$p = \frac{RT_e}{[(U/C) - \eta] (1 + \tau)^\gamma} \left[1 + \frac{C}{6W} - \frac{\alpha \gamma \tau}{1 + \tau} - \frac{C}{W} \frac{\xi^2}{2(1 + \tau)^2} - \left(\frac{\gamma}{2 - \gamma} \right) \frac{C}{W} \frac{RT_e}{V_0^2} \left\{ \frac{2 - \gamma}{\gamma - 1} + \frac{1}{1 + \tau} - \frac{1}{(\gamma - 1)(1 + \tau)^{\gamma - 1}} \right\} \right], \quad (24)$$

$$\frac{1}{\rho} = \left(\frac{U}{C} - \eta \right) (1 + \tau) \left[1 - \frac{C}{6\gamma W} + \alpha + \frac{C}{W} \frac{\xi^2}{2\gamma(1 + \tau)^2} + \frac{C}{W} \frac{RT_e}{(2 - \gamma) V_0^2} \left\{ \frac{2 - \gamma}{\gamma - 1} + \frac{1}{1 + \tau} - \frac{1}{(\gamma - 1)(1 + \tau)^{\gamma - 1}} \right\} \right]. \quad (25)$$

The terms in C/W and α stay small for all time, so the neglect of higher powers is almost certainly correct at all times and not merely near $\tau = 0$.

Dr Clemmow has shown me how this solution can be derived by a neater method based on the same order of approximation.

7. MOTION OF THE INNER BOUNDARY OF THE RAREFACTION WAVE

The velocity of sound

$$c = \frac{(\gamma RT_e)^{\frac{1}{2}}}{(1 + \tau)^{\frac{1}{2}(\gamma - 1)}} \left[1 + \alpha \left\{ 1 - \frac{(\gamma + 1)\tau}{2(1 + \tau)} \right\} + \frac{C}{W} \frac{RT_e}{V_0^2} \left\{ \frac{1}{2(2 - \gamma)(1 + \tau)^{\gamma - 1}} - \frac{1}{2} - \frac{(\gamma - 1)}{2(2 - \gamma)(1 + \tau)} \right\} \right]. \quad (26)$$

Here the terms in α^2 and $(C/W)^2$ have been neglected as usual, and also in this case a pair of terms proportional to C/W , whose sum is at most $C/30W$.

Substituting from (26) and (23) into (22), it is found that the boundary of the rarefaction wave satisfies the equation

$$\frac{dZ}{dt} - \frac{Z}{1+\tau} \left[1 + \frac{C}{W} \frac{RT_e}{(2-\gamma)V_0^2} \left\{ (1+\tau)^{1-\gamma} - (1+\tau)^{-1} \right\} \right] + \frac{(\gamma RT_e)^{\frac{1}{2}}}{V_0(1+\tau)^{\frac{1}{2}(\gamma-1)}} \times \left[1 + \alpha \left\{ 1 - \frac{(\gamma+1)\tau}{2(1+\tau)} \right\} + \frac{C}{W} \frac{RT_e}{V_0^2} \left\{ \frac{1}{2(2-\gamma)(1+\tau)^{\gamma-1}} - \frac{1}{2} - \frac{(\gamma-1)}{2(2-\gamma)(1+\tau)} \right\} \right] = 0, \quad (27)$$

of which the general solution is

$$\begin{aligned} & \frac{Z}{1+\tau} \left[1 + \frac{C}{W} \frac{RT_e}{(2-\gamma)V_0^2} \left\{ \frac{1}{(\gamma-1)(1+\tau)^{\gamma-1}} - \frac{1}{1+\tau} \right\} \right] \\ &= \frac{1}{1+\tau_0} \left[1 + \frac{C}{W} \frac{RT_e}{(2-\gamma)V_0^2} \left\{ \frac{1}{(\gamma-1)(1+\tau_0)^{\gamma-1}} - \frac{1}{1+\tau_0} \right\} \right] + \frac{2(\gamma RT_e)^{\frac{1}{2}}}{(\gamma-1)V_0(1+\tau)^{\frac{1}{2}(\gamma-1)}} \\ & \times \left[1 - \frac{\alpha(\gamma-1)\tau}{2(1+\tau)} + \frac{C}{W} \frac{RT_e}{V_0^2} \left\{ \frac{\gamma+1}{6(2-\gamma)(\gamma-1)(1+\tau)^{\gamma-1}} - \frac{1}{2} - \frac{(\gamma-1)}{2(2-\gamma)(1+\tau)} \right\} \right] \\ & - \frac{2(\gamma RT_e)^{\frac{1}{2}}}{(\gamma-1)V_0(1+\tau_0)^{\frac{1}{2}(\gamma-1)}} \left[1 - \frac{\alpha(\gamma-1)\tau_0}{2(1+\tau_0)} + \frac{C}{W} \frac{RT_e}{V_0^2} \right. \\ & \quad \times \left. \left\{ \frac{\gamma+1}{6(2-\gamma)(\gamma-1)(1+\tau_0)^{\gamma-1}} - \frac{1}{2} - \frac{(\gamma-1)}{2(2-\gamma)(1+\tau_0)} \right\} \right]. \quad (28) \end{aligned}$$

The new constant τ_0 is the value of τ when $Z = 1$, that is, when the boundary of the rarefaction starts from the muzzle. There are two cases to be considered.

Case I. Shot velocity less than local velocity of sound. The 'velocity of sound' here refers to the velocity in the gas just behind the shot at the instant of shot ejection. Case I is much the more common, and the condition for it to occur is, from (26),

$$V_0 < (\gamma RT_e)^{\frac{1}{2}} \left[1 + \frac{\eta}{(U/C) - \eta} \right]. \quad (29)$$

In this case the rarefaction wave moves in from the muzzle immediately after shot ejection, and therefore $\tau_0 = 0$. This simplifies (28) sufficiently to justify repeating the solution for this particular case:

$$\begin{aligned} & \frac{Z}{1+\tau} \left[1 + \frac{C}{W} \frac{RT_e}{(2-\gamma)V_0^2} \left\{ \frac{1}{(\gamma-1)(1+\tau)^{\gamma-1}} - \frac{1}{1+\tau} \right\} \right] \\ &= 1 + \frac{C}{W} \frac{RT_e}{(2-\gamma)V_0^2} - \frac{2(\gamma RT_e)^{\frac{1}{2}}}{(\gamma-1)V_0} \left[1 + \frac{C}{W} \frac{RT_e}{3(\gamma-1)V_0^2} \right] + \frac{2(\gamma RT_e)^{\frac{1}{2}}}{(\gamma-1)V_0(1+\tau)^{\frac{1}{2}(\gamma-1)}} \\ & \times \left[1 - \frac{\alpha(\gamma-1)\tau}{2(1+\tau)} + \frac{C}{W} \frac{RT_e}{V_0^2} \left\{ \frac{\gamma+1}{6(2-\gamma)(\gamma-1)(1+\tau)^{\gamma-1}} - \frac{1}{2} - \frac{(\gamma-1)}{2(2-\gamma)(1+\tau)} \right\} \right]. \quad (30) \end{aligned}$$

Case II. Muzzle velocity greater than local velocity of sound, that is

$$V_0 > (\gamma RT_c)^{\frac{1}{2}} \left[1 + \frac{\eta}{(U/C) - \eta} \right]. \quad (31)$$

This occurs for guns of very high muzzle velocity, of the order of 3500 ft./sec. Just after shot ejection the gas velocity at the muzzle is greater than the local velocity of sound, and hence the conditions inside the gun are completely independent of what is happening outside. The gas velocity at the muzzle soon falls to the local velocity of sound, and at this time the rarefaction wave starts from the muzzle towards the breech. The value of τ_0 is given by $v = c$ for $\xi = 1$ and $\tau = \tau_0$, and can be found numerically from (23) and (26) in any particular case. An explicit formula for τ_0 can be given, but it is complicated and not illuminating. Having found τ_0 numerically, it can be substituted in (28) to give the relation between Z and τ .

The solution for case II is valid only if the initial conditions (5)–(9) are true. It seems possible that when the velocity of the projectile exceeds the velocity of sound in the gas behind it the distribution in the barrel may differ from the simple result of (5)–(9).

8. COMPARISON WITH AN EXPERIMENTAL BREECH-PRESSURE RECORD

In applications, most of the parameters can be found without difficulty. RT_c can be calculated from Resal's equation, the energy equation of internal ballistics, using the calculated kinetic energy of shot and propellant gases, the thermochemical properties of the propellant, and an estimate of the heat loss to the bore. In the present example, the experimental breech pressure at shot ejection is known, which supplies RT_c immediately.

Consider an example for which $C/W = 0.474$, $\alpha = 0.197$, $V_0 = 3130$ ft./sec., and $\gamma = 1.275$ over the relevant temperature range. The breech pressure as the shot leaves is about $9\frac{1}{2}$ tons/sq.in. By (21) and (6) the velocity of sound just behind the shot as it leaves the muzzle is about 3640 ft./sec.; hence this is an example of case I, and the rarefaction wave starts from the muzzle at $\tau_0 = 0$. From (30) Z can be calculated, the position of the rarefaction boundary, at any time. The result is shown in figure 2. The boundary starts from the muzzle at a fairly low speed, of order 500 ft./sec., and accelerates as it approaches the breech, chiefly owing to the lower gas velocities in this region. The wave reaches the breech after 7 msec., which is roughly equal to the time taken for the shot to travel through the gun.

Figure 3 shows the breech pressure for times after shot ejection. The results of the present work are in good agreement with the experimental curve. The error in the pressure at any time is best stated as a fraction of the pressure drop from ejection up to that time. This error is always less than 5 %. It is interesting that the theoretical curve lies below the experimental results. The error is due to (i) neglect of terms of order $(C/W)^2$ and α^2 ; (ii) assumption that the cross-section is constant along the gun; (iii) neglect of heat loss to the barrel after the shot has left. It seems

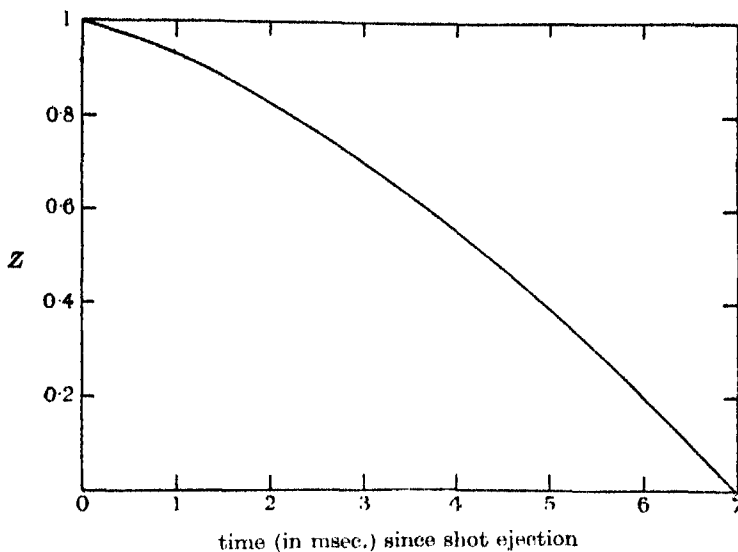


FIGURE 2. Motion of the rarefaction boundary.

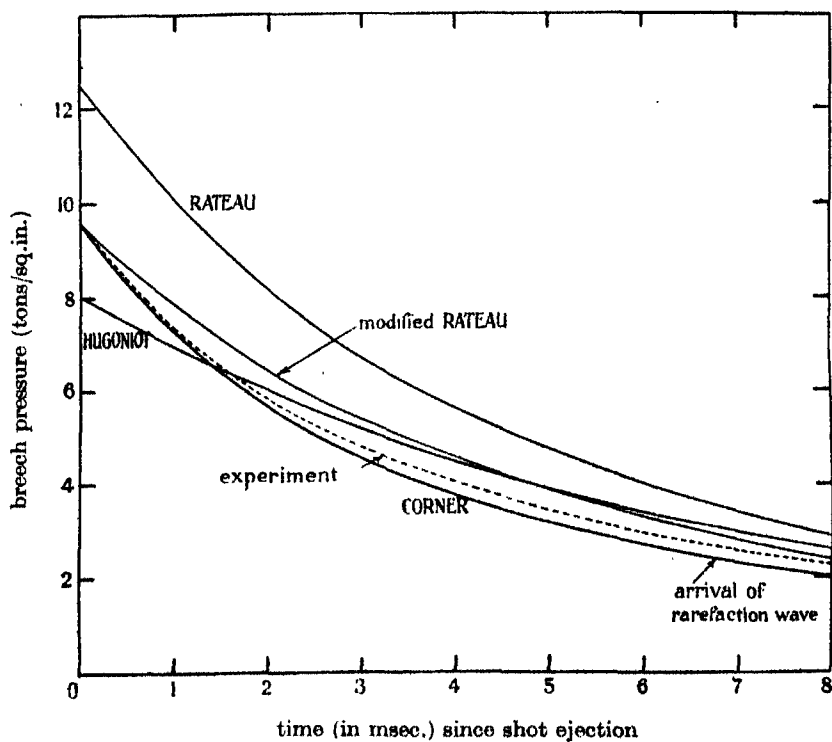


FIGURE 3. Breech pressure after shot ejection.

obvious that (iii) must cause the theoretical pressures to be higher than the true ones; therefore (i) and (ii) together must be responsible for lowering the theoretical pressure. The agreement within 5 % arises by partial compensation of opposing errors, but is nevertheless very satisfactory.

The results of some other theories are also shown on figure 3. The curve marked 'Rateau' was calculated with his correction for the initial motion of the gas (equations (10), (11), (12) and (13)). The 'modified Rateau' results (equations (14) and (15)) without this correction are seen to be in much better agreement with experiment. Hugoniot's formulae (16) and (17) are calculated by omitting all covolume terms, which accounts for the pressure lying considerably below the experimental values at the start. This is partly offset by the theoretical pressure being too big later.

The formulae for the breech pressure given in this paper hold until the rarefaction wave reaches the breech, which occurs after 7 msec. An exact treatment for later times would be difficult, and fortunately it is usually unnecessary. It is sufficient to complete the post-ejection period by one of the other theories. This has been done in figure 3 by using Hugoniot's formulae, with time measured from the instant at which the rarefaction wave first reaches the breech. It can be seen that in this case there is no appreciable kink in the pressure/time curve at this point. There does not appear to be any discontinuity of slope in the experimental results either. It is possible that some observable discontinuity of slope might occur if the wave were to reach the breech while the pressure was still high. This would imply a low-velocity gun with fairly high pressure as the shot leaves.

In many applications the important quantity to be derived from breech pressures after ejection is the *total* momentum caused by the gas pressure on the breech. The curves in figure 3, labelled 'modified Rateau' and 'Hugoniot', differ considerably in shape, while their total momenta in the post-ejection period are only 1 % apart. For certain calculations they are effectively identical.

Some momenta are therefore listed to show the differences between various theories. The momentum acquired from shot start to shot ejection

$$= (W + \frac{1}{2}C) V_0 = 2045 \text{ lb. wt.} \times \text{sec.}$$

TABLE 1. SUMMARY OF MOMENTA FROM BREECH PRESSURE (LB. WT. \times SEC.)

	experi- ment	present paper	Rateau	modified Rateau	Hugoniot
first 8.8 msec. after ejection	629	596	862	686	653
error (%)	—	- 5	+ 37	+ 11	+ 4
after first 8.8 msec.	—	261	335	281	322
total momenta	—	2902	3242	3011	3020

The post-ejection period contributes about 32 % of the total momentum. It is clear that the 'modified Rateau' results are practically as accurate as those of the present theory, and omission of the covolume (which gives the Hugoniot formulae) alters the results by a negligible amount. The latter formulae are much the simplest.

This was the reason why a Hugoniot curve was used for the period after the rarefaction wave had reached the breech.

It is possible to formulate a simple criterion to show which guns are particularly suitable for muzzle brakes. An obvious criterion is the ratio

$$\frac{\text{momentum from gas pressure after shot ejection in bare gun}}{\text{momentum from gas pressure up to shot ejection in bare gun'}}$$

which by integration of Hugoniot's formulae is proportional to $\frac{C(RT_e)^{\frac{1}{2}}}{(W + \frac{1}{2}C)\bar{V}_0}$. Let T_0 be the temperature of uncooled explosion of the propellant. Since T_e/T_0 is much the same for all guns,

$$\beta = \frac{C(RT_0)^{\frac{1}{2}}}{(W + \frac{1}{2}C)\bar{V}_0} \quad (32)$$

may be used as a criterion of suitability. The larger β , the more effective will be a muzzle brake of given geometrical design; it should be noticed, however, that β is not a property of the gun alone but depends also on the weight and type of propellant used.

A point of some interest is the extent to which β can be increased by increase of C/W . Langweiler (1938) has published muzzle velocities of a rifle with values of C/W from 0.86 to 44, and the same copper crusher pressures in all cases; actually the charge weight C was kept constant while the shot weight W was reduced to smaller and smaller values. In using such extreme values of C/W it must be remembered that β is to be multiplied by a factor $(1 + C/12W)$, which has been neglected in deriving (32), since it is so close to 1 for all ordinary guns; moreover, the standard corrections for gas inertia (Lagrange's ballistic problem) are not correct for such values of C/W , though Langweiler's results themselves show that the errors are not very large in his case. With these provisos in mind, it can be said that the criterion corresponding to β increases steadily with C/W , at constant crusher pressure.

Details of the velocity and pressure distributions along the gun will be given at the end of the next section, after an approximate treatment of the rarefaction zone.

9. THE RAREFACTION ZONE; MUZZLE-BRAKE CALCULATIONS

Consider first case I (shot velocity ^{less?} greater than local velocity of sound), in which the rarefaction starts back from the muzzle as soon as the shot leaves. Z , the position reached by the rarefaction boundary, can be calculated by (30); substitution in (23) and (25) gives the density and gas velocity at this boundary. These are denoted by $\rho_1(\tau)$ and $v_1(\tau)$. Let $\rho_2(\tau)$, $v_2(\tau)$ be the corresponding quantities at the muzzle; these are as yet unknown.

While $Z > 0$, it is possible to carry out the following approximate treatment of the motion inside the rarefaction zone. Assume a particular form for the way the density varies along the zone at any time, in terms of its values at the ends: $\rho_1(\tau)$

at the boundary $\xi = Z$, and $\rho_2(\tau)$ at the muzzle. The latter is a function not yet determined. Then take the equation of continuity

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0, \quad (33)$$

and integrate with respect to x , from the rarefaction boundary to the muzzle. Returning to the variables $\tau = V_0 St/U$ and $\xi = xS/U$, the result is

$$\rho_2 v_2 - \rho_1 v_1 = - \int_Z^1 \left(\frac{\partial \rho}{\partial \tau} \right) d\xi. \quad (34)$$

This gives $v_2(\tau)$. Also from $\rho_2(\tau)$, $c_2(\tau)$ is known, the velocity of sound at the muzzle. When $\tau = 0$, $c_2 > v_2$, for this is the definition of case I, which is being considered. In steady flow through a nozzle, $c = v$ at the throat, provided the pressure at entry is more than about twice the pressure in the region into which the nozzle is discharging. It is evident that v_2 must rapidly approach c_2 ; this condition decides the form of $\rho_2(\tau)$ and so the values of density and velocity throughout the rarefaction zone. For the prediction of muzzle-brake thrusts $\rho_2 v_2^2$ is needed, which is also obtained by this method.

The accuracy of this treatment depends on guessing a good form for the density variation in the rarefaction zone. The form chosen must be mathematically convenient also. A suitable assumption appears to be that the density is a linear function of distance in the rarefaction zone, so that

$$\rho(\xi, \tau) = \frac{(1 - \xi)\rho_1(\tau) + (\xi - Z)\rho_2(\tau)}{1 - Z}. \quad (35)$$

$$\text{From (22),} \quad V_0 \frac{dZ}{d\tau} = v_1 - c_1. \quad (36)$$

$$\text{Introduce a ratio} \quad \theta(\tau) = \rho_1(\tau)/\rho_2(\tau). \quad (37)$$

The equation (34) then leads to

$$v_2 = \theta v_1 - \frac{1}{2} V_0 (1 - Z) \left\{ (1 + \theta) \frac{d}{d\tau} (\ln \rho_1) - \frac{d}{d\tau} \ln \theta \right\} - \frac{1}{2} (c_1 - v_1) (1 - \theta), \quad (38)$$

$$\text{while} \quad c_2 = \frac{(\gamma RT_c)^{\frac{1}{2}}}{(1 - \eta \rho_2)^{\frac{1}{2}}} \left(\frac{(U/C) - \eta}{(1/\rho_2) - \eta} \right)^{\frac{1}{2}(\gamma-1)}. \quad (39)$$

It has been noticed that $\theta(\tau)$ must be chosen to make $c_2 - v_2$ rapidly approach zero from its initial positive value. It does not seem easy to do this in the general case, so take as an example the firing mentioned in § 8. Assume that

$$\theta = (1 + \tau)^n. \quad (40)$$

The behaviour of c_2 and v_2 (up to the time the rarefaction wave reaches the breach) shows that n must be taken to be 0.6. A more general form such as $\theta = (1 + m\tau)^n$ would be preferable, but it would be much more laborious to decide the best pair of

values of m and n ; moreover, c_2 is not very sensitive to θ , and hence the 'best possible' v_2 would probably not be much altered by the use of the extra parameter m .

The value of $n = 0.6$ applies to this example only, though n can be expected to have the same order of magnitude for all guns.

Figure 4 shows the theoretical pressure distribution along the barrel at various times after shot ejection. The values in the rarefaction zone were calculated from the treatment of this section with $n = 0.6$, and those closer to the breech by equation (24). It can be seen how the pressure drop down the barrel, which is about 2 tons/sq.in. when the shot is at the muzzle, is rapidly damped thereafter. There is, of course, a rapid drop of pressure near the muzzle, in the rarefaction wave; this drop attains about $1\frac{1}{2}$ tons/sq.in. in 2 msec., and stays practically constant for later times.

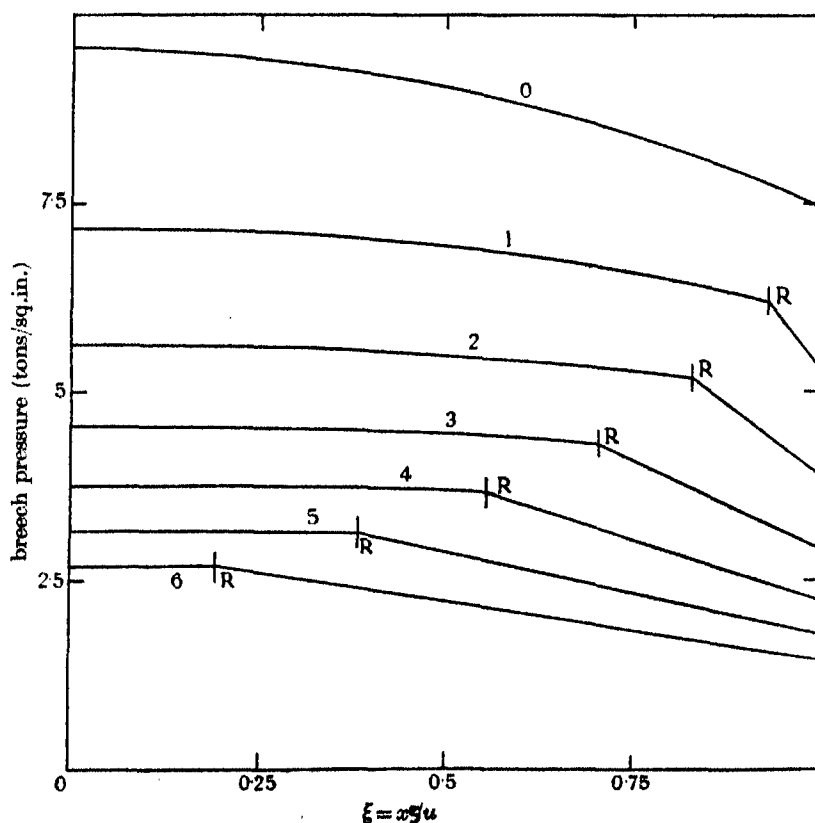


FIGURE 4. Theoretical pressure distributions at various times. R denotes boundary of rarefaction wave. Numbers on curves denote time (in msec.) after shot ejection.

In figure 5 are plotted local gas temperatures, calculated from coexisting values of pressure and density. The gas temperature at the muzzle drops by 500°C during the first 6 msec. The temperature drop in the rarefaction zone reaches nearly 200°C .

Figure 6 shows the local velocity of sound as a function of position and time. This was calculated from the density distributions. The gas velocities are also shown. To get the velocity at a point ξ in the rarefaction zone it is necessary to integrate (33) from the boundary Z to ξ . The rather complicated formulae is not given here. The velocities computed in this way cross in a confusing manner in figure 6. Their general trend is sufficiently clear, as is the fact that the gas velocity at the muzzle does indeed approach the local velocity of sound more closely as time goes on.

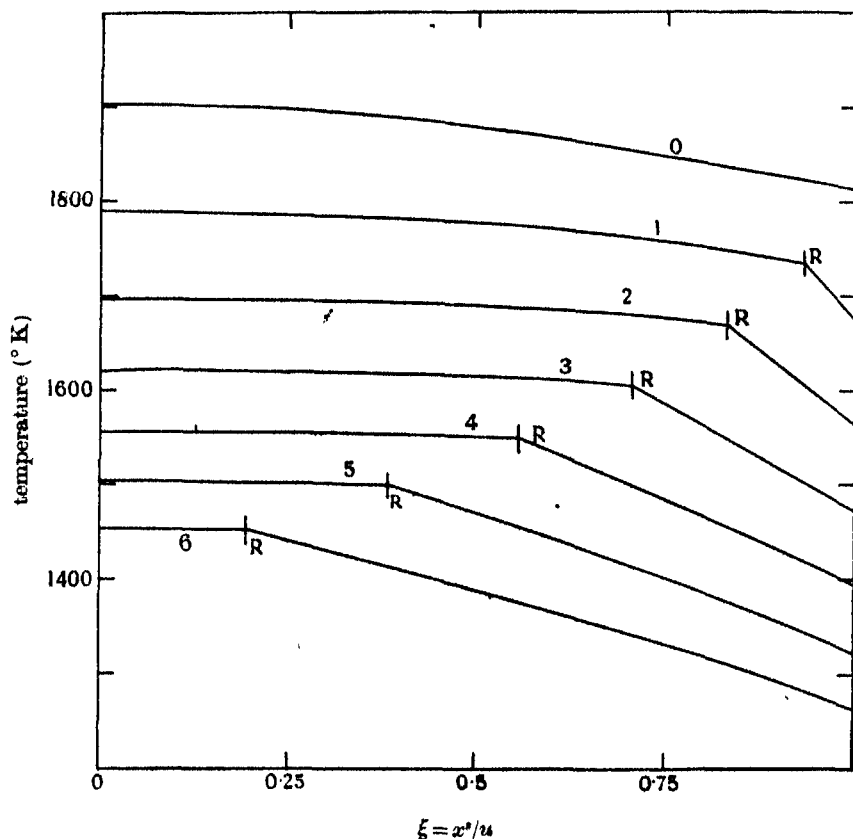


FIGURE 5. Local gas temperatures at various times. R denotes boundary of rarefaction wave. Numbers on curves denote time (in msec.) after shot ejection.

In theoretical prediction of muzzle-brake thrusts it is necessary to know the momentum of the gas passing the muzzle section and entering the brake. The momentum per second is $S\rho_g v_g^2$. Table 2 shows the results of the present work (with $n = 0.6$), the 'modified Rateau' theory with covolume, and the Hugoniot formulae. The first two theories differ by 25 % at the start and 20 % (in the opposite direction) after 6 msec. The significant measure of difference is, however, the total momentum of the gas crossing the muzzle section. This is only 1 % different in these two cases. The work of the present section applies only to the time before the rarefaction wave

reaches the breech. A theory of the conditions inside the gun at a later time will not be attempted. The modified Rateau formulae may be accepted as applicable at all times to within the accuracy needed for calculation of muzzle-brake thrusts, which include rather uncertain corrections for friction losses in the brake itself.

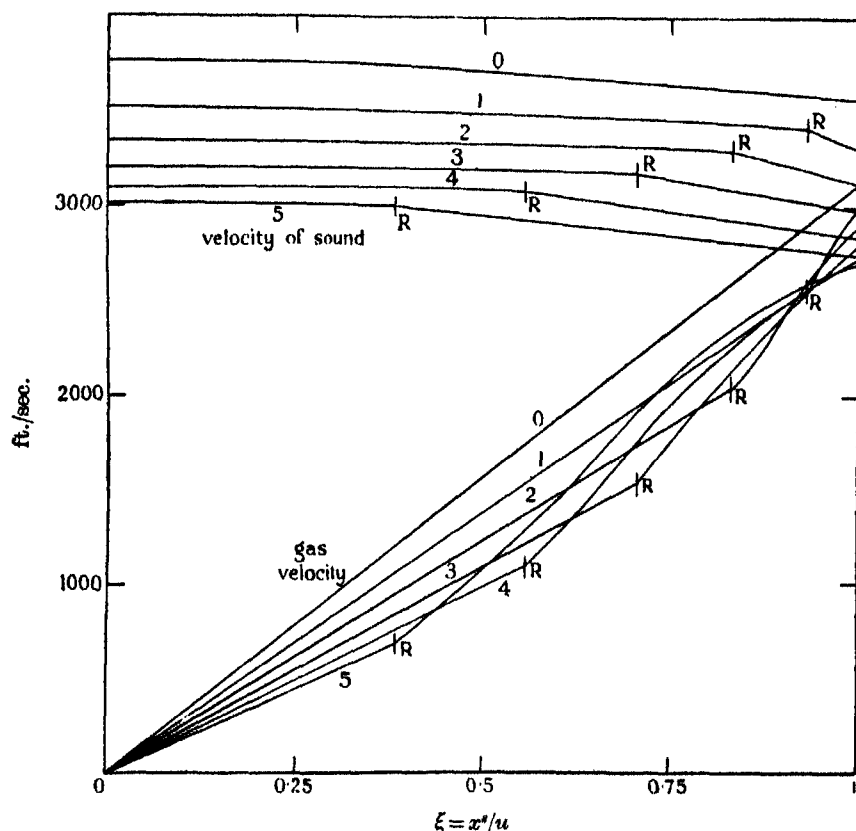


FIGURE 6. Local velocities at various times. *R* denotes boundary of rarefaction wave. Numbers on curves denote time (in msec.) after shot ejection.

TABLE 2. RATE OF OUTFLOW BY VARIOUS THEORIES

time from ejection, in msec.	0	1	2	3	4	5	6
$S\rho_1 v_1^2$, present theory, in tons wt.	64.1	46.5	34.5	26.4	20.9	17.1	14.3
'modified Rateau'	50.8	41.3	34.0	28.3	23.8	20.0	17.0
Hugoniot	38.3	33.0	28.5	24.6	21.4	18.6	15.8

It is true, of course, that while the projectile is passing through the brake this is acting with higher efficiency than later. Therefore the total momentum produced by the muzzle brake will be rather more than 1 % different in these two cases. This initial period lasts about $\frac{1}{4}$ msec. in the present example, and a rough calculation shows that the difference between the two theories cannot be more than 3 % for any

likely brake design. However, the stresses on the material are probably greatest in the period while the shot is passing through the brake, and if calculations of strength are to be derived from the modified Rateau formulae, suitable corrections must be applied.

By omitting the covolume terms one passes from the Rateau to the Hugoniot formulae. The total momentum of the gas is about 15 % too small on this theory. The error arises almost equally from two points: the under-estimate of the rate of loss of gas, and too low a velocity of sound when the covolume is omitted. The failure of the Hugoniot theory at this stage is unfortunate, since it appears to be adequate for the calculation of breech pressures, and it is much easier to use than the Rateau formulae.

Now consider case II. As the shot passes the muzzle the gas velocity just behind the shell is greater than the local velocity of sound. The solution inside the gun is determined by the initial conditions inside the gun only up to the time at which the gas velocity at the muzzle has fallen to the local velocity of sound. Until this time the solution of § 6 is valid everywhere inside the gun. From this solution it is seen that the gas velocity at the muzzle decreases much faster than the local velocity of sound, so it is no long time before the rarefaction wave starts in from the muzzle. The conditions inside the rarefaction wave can be treated in the same approximate way as for case I. The only difference is that one can impose the rigid condition $c_2 = v_2$ for $\tau \geq \tau_0$, and hence by the analogues of equations (38) and (39) one can obtain a differential equation for ρ_2 (the density at the muzzle), which is a function of τ .

In case I there is a more qualitative condition to satisfy: v must approach c , both calculated at the muzzle; this can best be examined by using a trial form for ρ_2 , which was taken to be $(1 + \tau)^{-n} \rho_1$. In case II it is not so necessary to take a trial form for ρ_2 but it is still very convenient. The exact differential equation for ρ_2 is a complicated non-linear type.

An example of case II could therefore be treated in much the same way as the more normal case I, and the matter need not be stressed further here.

I am indebted to Dr C. A. Clemmow, Mr V. de L. Dear and Dr F. Smithies for helpful discussions on the mathematical and practical aspects of this subject, to the Chief Scientific Officer, Ministry of Supply, for permission to publish this paper, and to Professor N. F. Mott for his interest in this work.

REFERENCES

- Hugoniot, A. 1886 *C.R. Acad. Sci., Paris*, **103**, 1002, 1178.
 Langweiler, H. 1938 *Z. tech. Phys.* **19**, 416.
 Love, A. E. H. & Pidduck, F. B. 1921 *Phil. Trans. A*, **222**, 167.
 Rateau, A. 1919 *C.R. Acad. Sci., Paris*, **168**, 330, 435, 581.
 Rateau, A. 1932 *Mémor. Artill. fr.* **11**, 5.

The theory of wedge indentation of ductile materials

BY R. HILL, E. H. LEE AND S. J. TUPPER

Branch for Theoretical Research, Armament Research Department, Ministry of Supply

(Communicated by N. F. Mott, F.R.S.—Received 28 August 1945)

[Plate 4]

The theoretical solution for the deformation produced by a rigid frictionless wedge penetrating a plastic material is presented in this paper. The solution determines the form of the lip and the deformation in the material squeezed out towards the surface. The variation with wedge angle of the force required for penetration is determined in terms of the yield stress in a tensile test.

The deformation of a grid of squares on a plane normal to the axis of the wedge is shown in figure 9 for a 30° semi-angle wedge. Close agreement is obtained with experiments as shown by comparing figure 9 with the photographs in figure 10 (plate 4). Lead blocks were used with a grid of squares scribed before penetration.

This investigation is a first step towards the correlation of hardness test results with the deformation properties of a material under other conditions of stress, e.g. the tensile test. The average strain due to wedge indentation is shown to correspond to an equivalent reduction of area in a tensile test which increases with increasing wedge angle.

1. INTRODUCTION

The indentation hardness test has not yet been satisfactorily analyzed in detail on the basis of the mathematical theory of plasticity. A complete analysis would calculate the load in terms of the geometry of the indenter and the stress-strain relationships of the material under uniform stress distributions, and would follow the development of the lip or the sinking in. In the present paper we discuss the problem of indentation of a material by a lubricated wedge as a first step to a solution of the more complex three-dimensional problem of indentation by a cone or ball.

The distributions of stress at a given instant, neglecting the change in form of the surface, were worked out in the early twenties for various two-dimensional indenters (cf. Prandtl 1923), but the progress of the deformation with increasing penetration was not followed. In this report the complete history of the motion is followed in detail, and full account is taken of the continually altering position of the surface and of the spreading of the plastic flow.

The theory is compared with experiments* in which lead blocks were indented by steel wedges coated with vaseline. It is found that the theory satisfactorily predicts the deformation of a grid of squares ruled on a cross-section of the block.

* Experiments carried out under extra-mural contract to the Ministry of Supply by Dr J. F. Allen and Mr J. M. Speight, Cavendish Laboratory, Cambridge.

2. FORMULATION OF THE PROBLEM

The metal under test is considered to have a sharp yield-point and negligible rate of work-hardening for the range of strains encountered. These conditions are satisfied by some pre-strained materials and by annealed steels with a marked lower yield-point extension. The indentation into such metals produces the raised lip type of impression. Since the material under the indenter can escape toward the free surface and into the lip, the effect of the elastic compressibility of the surrounding material in absorbing the displaced volume will be small. It might be expected therefore that the volume of material above the original plane surface would be roughly equal to the volume of impression below the surface. Confirmation of this is given in experiments on hardened copper and mild steel, reported by Ichihara (1931), in which measurements were made of the displaced surface in a Brinell test. We shall accordingly neglect any density changes during the indenting. Since, moreover, the displacements of the material which has flowed plastically from under the indenter into the lip are large compared with purely elastic displacements, it will be legitimate to treat the material as not only incompressible, but rigid until the flow stress is reached. The solution is not expected to represent satisfactorily indentation into materials such as lead and plasticine which do not possess a well-defined yield-point and exhibit a marked viscous strain-rate effect, or any material which is highly compressible. It is found, however, somewhat surprisingly, that the strain configuration in lead found from experiments (described in a later section) does, in fact, agree with theoretical calculations.

3. THEORY OF TWO-DIMENSIONAL PLASTIC FLOW

In this section we derive the plastic-flow equations in a form suitable for application to our problem. Since we are considering indentation by a long wedge into a semi-infinite medium the flow is confined to planes perpendicular to the length of the wedge and the problem becomes one of plane strain.

It is most convenient to refer stresses and strains in the plane to orthogonal curvilinear co-ordinates such that the two orthogonal directions at a point coincide with the directions of maximum shear stress. These two sets of curves of maximum shear stress, or slip-lines, are denoted by the parameters α and β respectively. In general the slip-lines alter position as the flow continues. The positive directions of the slip-lines are arbitrarily fixed by requiring (α, β) to be a right-handed pair and the shear stress $\widehat{\alpha\beta}$ (denoted by K) to be positive (figure 1). Whether Mohr's or von Mises' criterion for plastic flow is adopted does not affect the subsequent analysis, since, as will be shown, the maximum shearing stress is constant in both cases. It is chiefly for this reason that the axes of reference have been so chosen.

Consideration of the transformation of stress components at a point shows that the directions across which the shear stress is a maximum are subjected to equal

normal stress, denoted here by $-p$. The directions of principal stress at each point are at an angle of $\frac{1}{2}\pi$ to the slip-lines, the magnitudes of the principal stresses being $-p \pm K$.

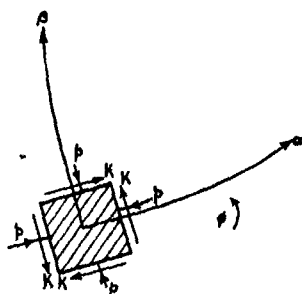


FIGURE 1

Now, if $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \tau_{xy}, \tau_{yz}, \tau_{zx}$ are the component stresses acting on an element with respect to a set of Cartesian axes, and if $d\epsilon_{xx}, d\epsilon_{yy}, d\epsilon_{zz}, d\epsilon_{xy}, d\epsilon_{yz}, d\epsilon_{zx}$ are the components of a possible natural strain increment under these stresses, then, since elastic strains are neglected, the following relations hold (Reuss 1930):

$$\left. \begin{aligned} d\epsilon_{xx} &= d\lambda \left\{ \sigma_{xx} - \frac{1}{2}(\sigma_{yy} + \sigma_{zz}) \right\}; & d\epsilon_{xy} &= 3d\lambda \tau_{xy}, \\ d\epsilon_{yy} &= d\lambda \left\{ \sigma_{yy} - \frac{1}{2}(\sigma_{zz} + \sigma_{xx}) \right\}; & d\epsilon_{yz} &= 3d\lambda \tau_{yz}, \\ d\epsilon_{zz} &= d\lambda \left\{ \sigma_{zz} - \frac{1}{2}(\sigma_{xx} + \sigma_{yy}) \right\}; & d\epsilon_{zx} &= 3d\lambda \tau_{zx}. \end{aligned} \right\} \quad (3.1)$$

$d\lambda$ is an unspecified parameter varying over the field. Let us take the z direction along the length of the wedge and the x, y axes to coincide with the α, β axes at the point under consideration. Then $\sigma_{xx} = \sigma_{yy} = -p$, $\tau_{xy} = K$, plane strain determines $d\epsilon_{zz} = d\epsilon_{xx} = d\epsilon_{yz} = 0$, and (3.1) reduces to

$$\sigma_{zz} = -p, \quad d\epsilon_{xx} = 0 = d\epsilon_{\beta\beta}. \quad (3.2)$$

The meaning of $d\epsilon_{xx}, d\epsilon_{\beta\beta}$ should be carefully noted: they are the components, in the momentary directions of the slip-lines, of the natural increment of strain of an element. $d\lambda$ can be found if required from $d\lambda = d\epsilon_{\alpha\beta}/3K$.

If Mohr's criterion of plastic flow is assumed (work-hardening being taken as negligible), σ_{zz} is the intermediate principal stress and K is the maximum shear stress in the plane of flow, so that $K = \frac{1}{2}Y$, where Y is the yield stress in tension. Von Mises' criterion is

$$(\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 + (\sigma_{xx} - \sigma_{yy})^2 + 6(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2) = 2Y^2,$$

and reduces to $K = Y/\sqrt{3}$.

Let ϕ (figure 1) be the angular displacement measured anti-clockwise of an α slip-line from some fixed direction. Then, denoting by ds_α, ds_β the elemental arc

lengths measured in the positive directions along the slip-lines, the radii of curvature R and S of the slip-lines may be defined by the relations

$$\frac{1}{R} = \frac{\partial \phi}{\partial s_\alpha}, \quad \frac{1}{S} = -\frac{\partial \phi}{\partial s_\beta}. \quad (3.3)$$

The equations of equilibrium, developed in the Appendix, are then

$$\frac{\partial p}{\partial s_\alpha} = -\frac{2K}{R}, \quad \frac{\partial p}{\partial s_\beta} = -\frac{2K}{S}. \quad (3.4)$$

It follows from (3.3) and (3.4) that

$$\left. \begin{aligned} p + 2K\phi &= \text{constant, along an } \alpha \text{ line,} \\ p - 2K\phi &= \text{constant, along a } \beta \text{ line.} \end{aligned} \right\} \quad (3.5)$$

These equations were originally derived by Hencky (1923). The elimination of p from (3.4) will give a relation determining the geometry of the slip-lines. This relation (developed in the Appendix) can be expressed as

$$\frac{\partial^2 \phi}{\partial \alpha \partial \beta} = 0, \quad (3.6)$$

having the solution

$$\phi = f(\alpha) + g(\beta),$$

where f and g are arbitrary functions.

This is equivalent to the statement that the angle between two slip-lines of the same family where they are cut by a slip-line of the other family is constant along their lengths. Any two families of orthogonal curves satisfying this condition constitute the field of slip-lines of a plastic mass in quasi-static flow under certain boundary conditions.

The boundary conditions will generally be in the form of given inclination ϕ of the slip-lines, and a relation between R and S on the boundary. The former is determined by the applied shearing stress, the latter by the applied normal forces. Thus, in the particular problem considered here, there are no applied forces on the free surface of the medium. The slip-lines therefore intersect this surface at an angle of $\frac{1}{2}\pi$. The space derivative of p along the boundary is, from (3.4),

$$\frac{\partial p}{\partial s} = \frac{1}{\sqrt{2}} \left(\frac{\partial p}{\partial s_\alpha} - \frac{\partial p}{\partial s_\beta} \right) = \frac{2K}{\sqrt{2}} \left(\frac{1}{S} - \frac{1}{R} \right).$$

Since p is constant and equal to K on the free surface, then

$$R = S. \quad (3.6a)$$

Since the slip-lines are characteristics of equations (3.6) two such fields can be joined together along a common slip-line without requiring the curvature of slip-lines of the other family to be continuous across this boundary. The simplest field of slip-lines consists of two sets of orthogonal straight lines. Another field consists of

arcs of concentric circles and radii. Both these fields are immediately seen to satisfy the geometrical interpretation of (3.6) given above. They are, in fact, all that is necessary to solve the problem.

As the penetration k of the wedge is increased (quasi-statically) by an amount Δk , the co-ordinates (referred to fixed Cartesian axes) of an element in the plastic region will increase by Δx , Δy , say. Denote the components of the vector

$$(\Delta x/\Delta k, \Delta y/\Delta k)_{\Delta k \rightarrow 0}$$

in the slip-line directions by (u, v) and let this vector be known as the 'velocity' of the element. Equation (3.2), governing the strain increment in an element, can be expressed in terms of the field velocities, as shown in the Appendix. Hence

$$\frac{\partial u}{\partial s_\alpha} = \frac{v}{R}, \quad \frac{\partial v}{\partial s_\beta} = \frac{u}{S}. \quad (3.7)$$

4. INDENTATION BY A LUBRICATED WEDGE

In the last section equations were established in a form suitable for the problem. Suppose that the wedge is perfectly lubricated. The plastic region is shown diagrammatically in figure 2 at a particular stage of the indenting. AC , $A'C'$ is the displaced free surface and $C'BC$ is the boundary between the material in plastic flow and the material still rigid. O is the point where the penetration began on the original free surface. The penetration $OB = k$ as already defined.

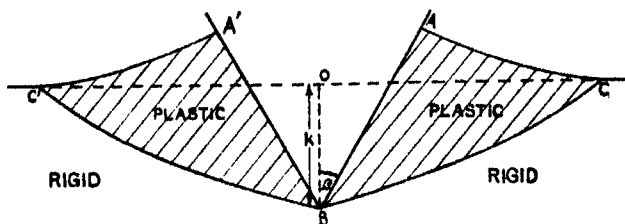


FIGURE 2

Since the surface AC is free from applied force, in particular applied tangential stress, it is a principal stress direction which the slip-lines intersect at an angle $\frac{1}{2}\pi$. They also intersect AB at $\frac{1}{2}\pi$, since perfect lubrication along AB is assumed and therefore no applied tangential stress. The velocity condition along AB (the meaning of velocity as used here being defined in § 3) is that the velocity of the material normal to the wedge is equal to the component of wedge velocity in the same direction. Since the velocity of penetration is unity this means that the component velocity of the material normal to the wedge is $\sin \beta$, where β = semi-angle of wedge. (It will always be clear whether β refers to the angle or to one set of slip-lines.)

The boundary BC has been taken through the tip of the wedge because, as will be shown, the solution constructed on this assumption satisfies all the boundary

During a small increment Δk in penetration, AC moves normal to itself through a distance $u\Delta k/\sqrt{2} = \sin \beta \Delta k$. If ON is the perpendicular from O to CA , the condition that the new configuration is similar to the original one is simply $\Delta k \sin \beta / \Delta k = ON/k$, i.e.

$$ON = k \sin \beta. \quad (4.2)$$

It is only necessary to impose the similarity condition on AC since, as has been seen, the whole stress and velocity distribution is uniquely determinate once the position of AC has been chosen. (4.2) gives another relation between h and α which, with (4.1), determines them both. (4.2) can be expressed, using the geometry of figure 3, as

$$h \cos \alpha = k(\sin \beta + \cos(\beta - \alpha)). \quad (4.3)$$

Eliminating h/k from (4.1) and (4.3) gives an equation between α and β

$$\cos(2\beta - \alpha) = \frac{\cos \alpha}{1 + \sin \alpha}. \quad (4.4)$$

The graph of α against β is shown in figure 8.

From (4.2) it can be seen that a circle centre O , radius ON , will touch AB and CA produced (figure 4). Furthermore, the triangles AOC , AOB have one side equal ($AC = AB$) and equal perpendiculars ON , OM , to those sides from the opposite vertices. Their areas are therefore equal. If OC cuts AB in L , this implies that triangles ALC , OLB are equal in area; in other words the volume of the lip is equal to the volume of displaced material, as it must be by the incompressibility assumption.

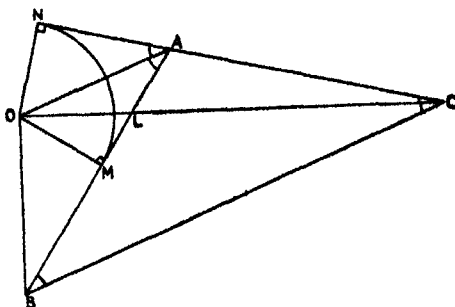


FIGURE 4

Now return to (3.5) to calculate the stress acting on the wedge. On the free surface $p = K$, since the principal stress normal to the surface $-p + K = 0$, and so, following an α slip-line from the surface to the wedge, $p = K(1 + 2\alpha)$ along the wedge. The pressure P acting on the wedge, i.e. the principal compression stress $p + K$, is therefore uniform and equal to $2K(1 + \alpha)$. P is plotted as a function of β in figure 8.

It is interesting to compare this variation of P with β for the two-dimensional problem with that found experimentally in the axially symmetric problem (Bishop, Hill & Mott 1945). In the latter the mean pressure P decreases with increasing β

but remains roughly constant for $\beta > 30^\circ$. For indentation with lubricated cones in hardened copper $P \sim 2.3Y$ in this range of β . If Von Mises' criterion is taken, giving $K = Y/\sqrt{3}$, the two-dimensional calculated value of P varies between $1.5Y$ and $3.0Y$, when β lies between 30 and 90° .

5. DEFORMATION OF AN ELEMENT DURING INDENTATION

In following the motion of a particular element directly, not only is its velocity influenced by its varying position in space, but the continuously expanding velocity field must also be taken into account. A direct determination of the history of the deformation of a particular element in these circumstances appears extremely complicated. However, one can use a simple transformation which overcomes this difficulty by following the motion of an element in an associated field, which is the actual field scaled so that the penetration is always unity. If \mathbf{r} denotes the actual position vector of the element with reference to O when the penetration is k , put $\mathbf{r} = k\mathbf{r}'$, so that \mathbf{r}' is the corresponding position vector on the unit diagram. Now $d\mathbf{r}/dk = \mathbf{v}$, the velocity of the element, and so

$$k \frac{d\mathbf{r}'}{dk} = \mathbf{v} - \mathbf{r}'. \quad (5.1)$$

It has been shown in the previous section that \mathbf{v} is a vector of constant length $\sqrt{2} \sin \beta$ for material in plastic flow; for material still rigid \mathbf{v} is, of course, zero. The point with position vector \mathbf{v} lies, therefore, on the arc HK of a circle, centre O , radius $\sqrt{2} \sin \beta$ in the unit diagram (figure 5). H corresponds to the velocity vector of all elements in region I, and OH is thus parallel to EC . K corresponds to elements in region III and OK is parallel to BD . An intermediate vector OP corresponds to elements in II lying on the radius AQ perpendicular to OP . This circle will be known as the focal circle and the points H, P, K as foci. (5.1) then expresses the fact that the velocity of an element on the unit diagram is directed toward the focus corresponding to its momentary position. When the element is still rigid its motion on the unit diagram is directed toward O . The trajectories in I and III are simply segments of straight lines directed toward H and K respectively; the trajectories in II are curved. To find the actual trajectory of a particular element in space it is only necessary to magnify the diagram at each stage by the appropriate value of k .

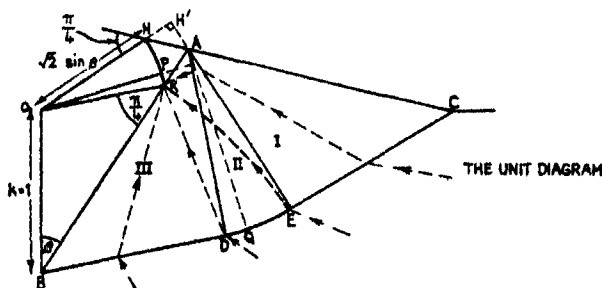
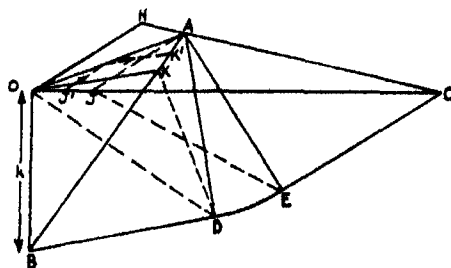


FIGURE 5



It is not possible to describe the motion of material initially inside $OJED$ in any simple way, other than has been done qualitatively on the unit diagram. The motion of an element J' formerly between O and J on the undisturbed surface is, however, easy to follow. The element first moves as a point of I parallel to EC until it reaches OA when it is overtaken by the wedge; thereafter it remains in contact with the wedge (though moving relatively to it) and moves as a point of III, parallel to BD , to K' .

For comparison with experiment it is convenient to calculate the deformation of a grid of squares ruled on a cross-section perpendicular to the wedge. It is necessary to calculate the final position, when the penetration is k , of a corner P of a square where initial position vector is \mathbf{r}_0 relative to O . Suppose the plastic boundary first reaches P when the penetration is k_0 ($< k$), and let $\mathbf{r}'_0 = \mathbf{r}_0/k_0$ be the position vector on the unit diagram at this stage. The point on the unit diagram now moves along the corresponding trajectory through \mathbf{r}'_0 which must be calculated by a step-by-step integration in region II. If s is the distance moved along the trajectory from \mathbf{r}'_0 when the penetration is k , and $f(s)$ is the focal distance, then from (5.1)

$$\log \frac{k}{k_0} = \int_0^s \frac{ds}{f(s)}. \quad (5.2)$$

This determines s in terms of k and so the position vector $\mathbf{r} = k\mathbf{r}'$ of the element in space. The integral in (5.2) has to be evaluated numerically in II. While the trajectory lies in I or III (not having crossed II) $f(s) = d - s$, where d is the distance from \mathbf{r}'_0 of the focus H or K respectively. (5.2) then simplifies to

$$\frac{k}{k_0} = \frac{d}{d - s}. \quad (5.3)$$

It is useful to notice that the same trajectory corresponds to all points initially on a radius through O . Thus the plastic boundary first reaches a point $\lambda\mathbf{r}_0$ when the penetration is λk_0 , so that $\mathbf{r}'_0 = \mathbf{r}_0/k_0$ as before. The final position s' on the trajectory when the penetration is k is found from

$$\log \frac{k}{\lambda k_0} = \int_0^{s'} \frac{ds}{f(s)}.$$

The same process of numerical evaluation of the integral serves therefore for all points $\lambda\mathbf{r}_0$.

Figure 9 shows the final configuration of a grid initially of squares with sides parallel and perpendicular to the surface. This will be discussed more fully in the next section.

6. COMPARISON WITH EXPERIMENT

To obtain a fair comparison of theory with experiment it is necessary to choose a material with a well-defined yield-point and negligible rate of work-hardening over the range of strains encountered. Mild steel is probably most satisfactory in these respects, but in practice it is more convenient to use a softer material, in this case lead.

Prior to indentation each lead block was cut in half and a grid of squares ruled on one section, the two sections being firmly held together during the indenting. The penetrations were carried out using steel wedges, with vaseline as the lubricant. The final configurations of the grids are shown in the photographs in figure 10 (plate 4).

The theoretical calculation of the final configuration with $\beta = 30^\circ$ is shown in figure 9. The squares lying completely in *OBD* or *JEC* (figure 6) become parallelograms in the final configuration; the intermediate stages of the deformation of these squares can be seen from those which are still partly rigid. The straining is most severe near the apex of the lip, where material originally on the free surface now forms the inner edge of the lip.

The existence of three distinct regions is clearly seen in the photographs. Moreover, both the extent and the manner of bending of the central region (containing material which has passed into or through II) agrees satisfactorily with experiment. This is most clearly seen from the widest angle. The assumption that the plastic-rigid boundary passes through the tip of the wedge is also confirmed, the slight discrepancies being attributable to the fact that the lead does not possess a well-defined yield-point. The volume of material in the lip in each case is roughly equal to, or somewhat less than, the volume of material displaced, so that the assumption of incompressible flow is not far from the truth even for such a compressible material as lead. This conclusion is restricted to the lubricated case, since there is scarcely any lip when no lubricant is used. There is apparently a little residual friction between the wedge and the lead responsible for the departure of the surface from strict linearity near the apex of the lip. In experiments with unlubricated wedges, it was found that the surface was quite rounded owing apparently to material being dragged along by the wedge. It is likely, however, that work-hardening also contributes to the curvature of the surface.

Although experiment substantially confirms the assumption that the plastic boundary passes through the wedge tip, a theoretical verification is also needed. This would entail a solution of the elastic stress equations in the material still rigid, using the distribution of forces already found on the plastic-rigid boundary. If the shear stress exceeded K the plastic region would need to be extended, though this would not necessarily mean additional flow since material could be in a condition of incipient flow but constrained by its relative incompressibility. Since the problem is two-dimensional the elastic stress solution is independent of elastic constants so that the results of experiments on a particular material, lead, have general significance.

7. THE AVERAGE STRAIN IN AN INDENTATION TEST

For the comparison of different testing methods it is important to be able to compare the amount of deformation to which a material is subjected in different tests. This is effected from a study of stress-strain curves under combined stresses of a material which work hardens. Defining the generalized stress $\bar{\sigma}$ by

$$\bar{\sigma} = \frac{1}{\sqrt{2}} \{ (\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 + 6(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2) \}^{\frac{1}{2}},$$

and the generalized plastic strain increment $\bar{d\epsilon}$ by

$$\bar{d\epsilon} = \frac{\sqrt{2}}{3} \{ (d\epsilon_{xx} - d\epsilon_{yy})^2 + (d\epsilon_{yy} - d\epsilon_{zz})^2 + (d\epsilon_{zz} - d\epsilon_{xx})^2 + \frac{3}{2}(d\epsilon_{xy}^2 + d\epsilon_{yz}^2 + d\epsilon_{zx}^2) \}^{\frac{1}{2}},$$

where $d\epsilon_{xx}$, $d\epsilon_{yy}$, etc., are increments of permanent plastic strain, the relationship between $\bar{\sigma}$ and $\bar{\epsilon} = \int \bar{d\epsilon}$ can be considered as a generalized stress-strain curve. It has been shown experimentally that a reasonable approximation to one generalized stress-strain curve is obtained for all types of loading under combined stresses, though most of the tests have been restricted to cases in which the principal stresses increase proportionately (cf. Ros & Eichinger 1926, 1928). Assuming that a unique $\bar{\sigma}$, $\bar{\epsilon}$ relationship obtains for all types of deformation, the deformation determined in the previous section for a wedge indentation test can be compared with the equivalent reduction of area in a tensile test. In particular, a check will be obtained on the assumption of zero work-hardening from the amount of hardening in a tensile test at the equivalent average strain corresponding to the solution given.

An interesting property of the $\bar{\sigma}$, $\bar{\epsilon}$ curve is that the total plastic work done per unit volume of unstressed material up to a given final equivalent deformation $\bar{\epsilon}$ is always the same, and is equal to the area $\int_0^{\bar{\epsilon}} \bar{\sigma} d\bar{\epsilon}$ under the generalized stress-strain curve. This can be seen by using (3.1) in the equation for the increment of plastic work per unit volume

$$dW = \sigma_{xx} d\epsilon_{xx} + \sigma_{yy} d\epsilon_{yy} + \sigma_{zz} d\epsilon_{zz} + \tau_{yz} d\epsilon_{yz} + \tau_{zx} d\epsilon_{zx} + \tau_{xy} d\epsilon_{xy},$$

giving $dW = d\lambda \bar{\sigma}^2$. But $\bar{d\epsilon} = d\lambda \bar{\sigma}$ from (3.1) and so $dW = \bar{\sigma} \bar{d\epsilon}$. Since the plastic volume change is zero

$$\int_0^{\bar{\epsilon}} \bar{\sigma} \bar{d\epsilon}$$

is the total work per unit volume of unstressed material.

In a tensile test the components of plastic natural strain are $(\epsilon, -\frac{1}{2}\epsilon, -\frac{1}{2}\epsilon)$, which determines

$$\bar{\epsilon} = \epsilon. \quad (7.1)$$

Also $\bar{\sigma} = \sigma$, so that the $\bar{\sigma}$, $\bar{\epsilon}$ curve coincides with the true-stress natural plastic strain curve in a tensile test. This is the reason for the choice of numerical factors in the definition of $\bar{\sigma}$ and $\bar{d\epsilon}$.

For the complicated strain distribution resulting from wedge indentation the final deformation of a particular element is obtained by following its motion throughout indenting and evaluating $\int \bar{d\epsilon}$. The final deformation of all elements which have not passed through region II is the same; the strain becomes infinitely large for the material initially immediately under the point of the wedge. Since the central region is small compared with the whole plastic region, the mean equivalent tensile strain can be taken as $\int \bar{d\epsilon}$ evaluated for an element in regions I or III. This is less justified for large values of β .

Since $\bar{d\epsilon}$ is invariant with respect to choice of axes, the value of $\int \bar{d\epsilon}$ is independent of the initial shape of a small element. It is convenient to take a square element

with a side parallel to BD (figure 6), and initially below OD so that the element always moves in III. The element deforms into a parallelogram (figure 7); there is also a translation of the element as a whole which is of no concern here. The element will not be a parallelogram during the intermediate stages of deformation, while part of it is still rigid. In evaluating $\int \bar{d}\epsilon$, however, it is immaterial by what process of straining the final configuration is reached so long as the particles of the element follow their actual paths at some continuously positive rate of strain. It is most convenient to consider the final configuration as being reached by simple shearing of the element. In a strain increment corresponding to a small increment of θ (figure 7),

$$\bar{d}\epsilon = \frac{d(\tan \theta)}{\sqrt{3}}, \quad (7.2)$$

and so the equivalent natural strain ϵ_T in a tensile test is

$$\epsilon_T = \int \bar{d}\epsilon = \frac{\tan \theta}{\sqrt{3}}. \quad (7.3)$$

The velocity of particles in III is $\sqrt{2} \sin \beta$, and the velocity of BD normal to itself is $\cos(\frac{1}{4}\pi - \beta)$. Thus

$$\tan \theta = \frac{\sqrt{2} \sin \beta}{\cos(\frac{1}{4}\pi - \beta)},$$

and

$$\epsilon_T = \sqrt{\frac{2}{3}} \sin \beta \sec(\frac{1}{4}\pi - \beta). \quad (7.4)$$

The equivalent percentage reduction in area is $100(1 - e^{-\epsilon_T})$; the maximum reduction is 68.5 % when $\beta = \frac{1}{2}\pi$. In the case worked out in detail here $\beta = \frac{1}{3}\pi$ and $\epsilon_T = 0.42$. It can now be seen by how much the solution, based on zero rate of work-hardening, needs modification when the actual stress-strain curve of a pre-strained metal is considered.

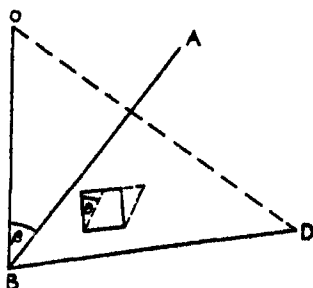


FIGURE 7

A true-stress natural-strain curve is given (Bishop *et al.* 1945) for copper pre-strained to have a yield-point of 17.5 tons/sq.in. At a strain of 0.42 the stress is 20 tons/sq.in. For a 0.15 C mild steel supposed cold-worked to a stress of 38 tons/sq.in. (MacGregor 1938), the stress at a strain of 0.42 is 48 tons/sq.in. The corresponding stress was 9.8 tons/sq.in. for pure aluminium cold-worked to a stress of 8.4 tons/sq.in. The major effect of work-hardening would presumably be to smooth

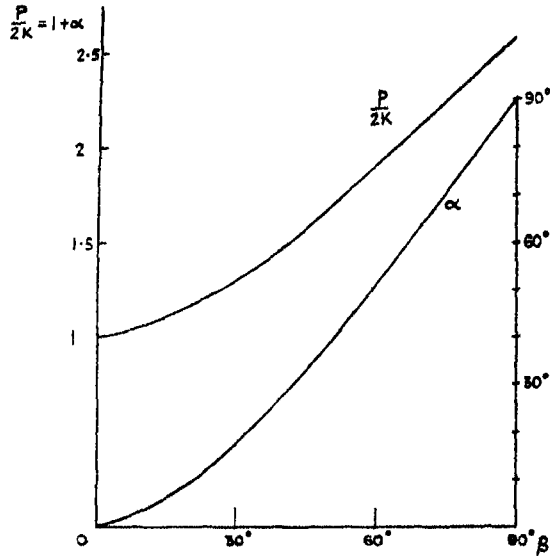


FIGURE 8. Variation of pressure P on wedge with semi-angle β .

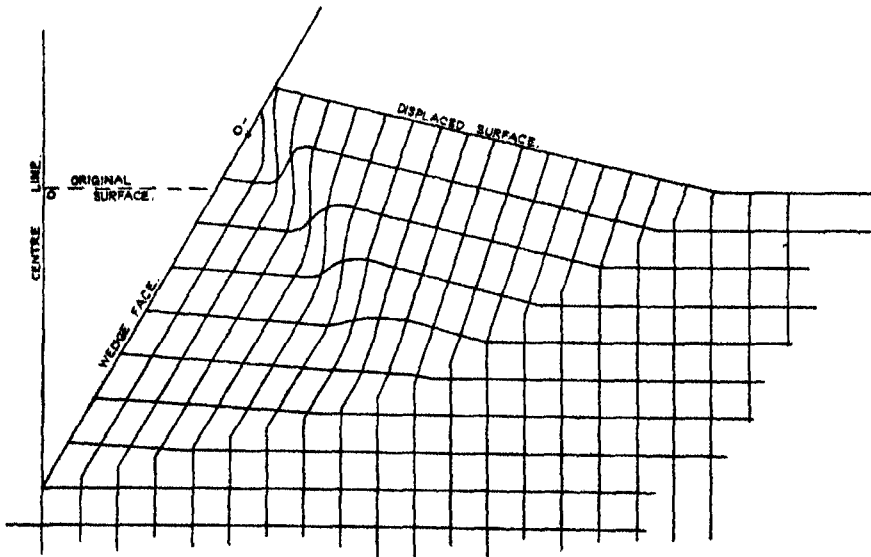


FIGURE 9. The deformation of a net of squares due to indentation.

out the discontinuity in shear strain at the plastic boundary. Since the strain in the plastic region is sensibly uniform the solution would apply approximately with a modified yield stress.

We are indebted to the Director-General of Scientific Research and Development, Ministry of Supply, for permission to publish this paper.

APPENDIX I

Plastic deformation relationships in terms of the slip-line field

The flow is expressed in terms of the curvilinear co-ordinates α, β which determine a point as the intersection of two slip-lines, one from each of the two orthogonal families α and β . If (x, y) are the Cartesian co-ordinates of a point, regarded as functions of α and β ,

$$dx = \frac{\partial x}{\partial \alpha} d\alpha + \frac{\partial x}{\partial \beta} d\beta, \quad dy = \frac{\partial y}{\partial \alpha} d\alpha + \frac{\partial y}{\partial \beta} d\beta.$$

Along an α line, $d\beta = 0$. Hence

$$ds_\alpha^2 = dx^2 + dy^2 = \left\{ \left(\frac{\partial x}{\partial \alpha} \right)^2 + \left(\frac{\partial y}{\partial \alpha} \right)^2 \right\} d\alpha^2,$$

i.e.

$$ds_\alpha = h_\alpha d\alpha, \quad (\text{A. 1})$$

where

$$h_\alpha^2 = \left(\frac{\partial x}{\partial \alpha} \right)^2 + \left(\frac{\partial y}{\partial \alpha} \right)^2.$$

Similarly

$$ds_\beta = h_\beta d\beta, \quad (\text{A. 2})$$

where

$$h_\beta^2 = \left(\frac{\partial x}{\partial \beta} \right)^2 + \left(\frac{\partial y}{\partial \beta} \right)^2.$$

From the geometry of figure (A. 1), representing an element bounded by adjacent slip-lines $\alpha, \alpha + d\alpha; \beta, \beta + d\beta$, it is found that

$$\frac{\partial \phi}{\partial \alpha} d\alpha = \frac{-\frac{\partial}{\partial \beta} (h_\alpha d\alpha) \cdot d\beta}{h_\beta d\beta},$$

i.e.

$$\frac{\partial \phi}{\partial \alpha} = -\frac{1}{h_\beta} \frac{\partial h_\alpha}{\partial \beta}. \quad (\text{A. 3})$$

Similarly

$$\frac{\partial \phi}{\partial \beta} = \frac{1}{h_\alpha} \frac{\partial h_\beta}{\partial \alpha}. \quad (\text{A. 4})$$

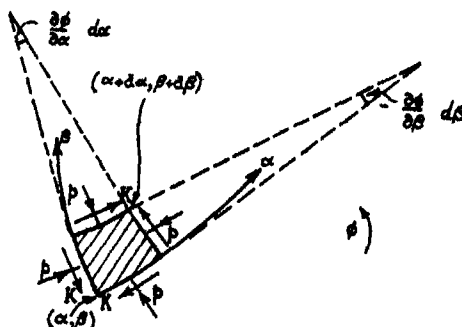


FIGURE A. 1

The element in figure A. 1 is in equilibrium under the stress components shown. Equating the resultant component of force in the α direction to zero, then

$$-\frac{\partial}{\partial \alpha}(ph_\beta)d\alpha d\beta + ph_\alpha \frac{\partial \phi}{\partial \beta} d\alpha d\beta - Kh_\beta \frac{\partial \phi}{\partial \alpha} d\alpha d\beta + \frac{\partial}{\partial \beta}(Kh_\alpha)d\alpha d\beta = 0.$$

Using (A. 3, 4) and the fact that K is constant, this becomes

$$\frac{\partial p}{\partial \alpha} + 2K \frac{\partial \phi}{\partial \alpha} = 0, \quad (\text{A. 5})$$

or, from (A. 1) and (3.3),
$$\frac{\partial p}{\partial s_\alpha} = -\frac{2K}{R}. \quad (3.4a)$$

Resolving in the β direction gives

$$\frac{\partial p}{\partial \beta} - 2K \frac{\partial \phi}{\partial \beta} = 0 \quad (\text{A. 6})$$

and
$$\frac{\partial p}{\partial s_\beta} = -\frac{2K}{S}. \quad (3.4b)$$

It should be noted that the arc-lengths s_α , s_β refer only to individual slip-lines and cannot be used as independent variables throughout the field. p can be eliminated from (A. 5) and (A. 6), giving

$$\frac{\partial^2 \phi}{\partial \alpha \partial \beta} = 0. \quad (3.6)$$

Differentiating the relation $R \frac{\partial \phi}{\partial \alpha} = h_\alpha$, then

$$\begin{aligned} \frac{\partial R}{\partial \beta} \frac{\partial \phi}{\partial \alpha} &= \frac{\partial h_\alpha}{\partial \beta} \quad \text{from (3.6),} \\ &= -h_\beta \frac{\partial \phi}{\partial \alpha} \quad \text{from (A. 3).} \end{aligned}$$

From (A. 2) this becomes
$$\frac{\partial R}{\partial s_\beta} = -1. \quad (\text{A. 7})$$

By a similar argument
$$\frac{\partial S}{\partial s_\alpha} = -1. \quad (\text{A. 8})$$

The element AB of length ds_α , originally coinciding with an α slip-line, moves to $A'B'$ during an increment of penetration dk . The variation in the u component of velocity, as shown in figure A. 2, results in a strain increment $\partial u / \partial s_\alpha dk$. In addition, the element moves towards the centre of curvature O by an amount vdk involving a compressive strain vdk/R . The resultant increment of strain, which must be zero from (3.2), gives

$$\frac{\partial u}{\partial s_\alpha} - \frac{v}{R} = 0. \quad (3.7a)$$

Similarly
$$\frac{\partial v}{\partial s_\beta} - \frac{u}{S} = 0. \quad (3.7b)$$

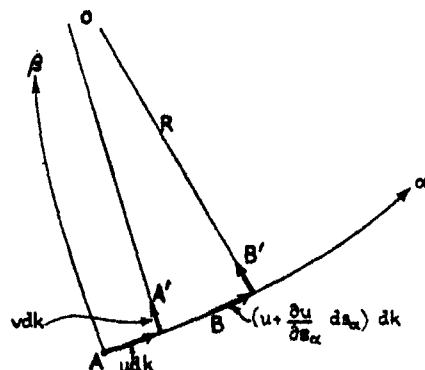


FIGURE A.2

REFERENCES

- Bishop, R. F., Hill, R. & Mott, N. F. 1945 *Proc. Phys. Soc.* **57**, 147.
 Hencky, H. 1923 *Z. angew. Math. Mech.* **3**, 4, 241-251.
 Ichihara, M. 1931 A contribution to Brinell hardness tests. *Tech. Rep. Tohoku Univ.* **10**, 25.
 MacGregor, C. W. 1938 *Differential area relations in the plastic state*. Timoshenko Anniversary Volume. Macmillan, U.S.A.
 Prandtl, L. 1923 *Z. angew. Math. Mech.* **3**, 6, 401.
 Reuss, A. 1930 *Z. angew. Math. Mech.* **10**, 3, 266.
 Ros & Eichinger 1926, 1928 *Versuche zur Klärung der Bruchgefahr*. Zurich.

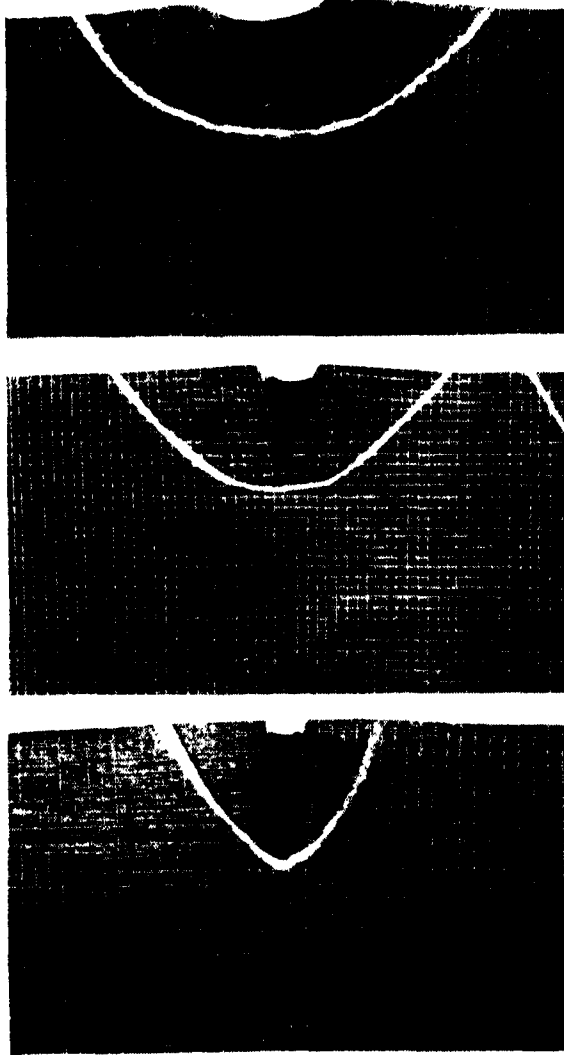


FIGURE 10. The thick white lines indicate very roughly the extent of the plastic flow.

The detonation of liquid explosives by gentle impact.

The effect of minute gas spaces

By F. P. BOWDEN, M. F. R. MULCAHY, R. G. VINES AND A. YOFFE, *Tribophysics Section, Council for Scientific and Industrial Research, University of Melbourne*

(Communicated by Sir David Rivett, F.R.S. and Sir Robert Robertson, F.R.S.—

Received 12 January—Read 9 May 1946)

An experimental study has been made of the physical conditions which occur during the impact of solids on liquid explosives. It is shown that the high impact sensitivity of liquid or gelatinous explosives is due to the entrapping of minute gas spaces during the impact. These tiny gas bubbles (the mass of bubble may be about 10^{-10} g. and its volume at atmospheric pressure about 10^{-7} c.c.) are heated by adiabatic compression and initiate the explosion. Bubbles of this size are, of course, difficult to detect, and they are readily entrapped under many conditions of laboratory experiment and practical operation. Because of the effect of the entrapped gas spaces the sensitivity of liquid explosives is profoundly affected by the shape of the impacting surfaces or by the distribution of the explosive. If, for example, one of the surfaces contains a small pin-hole or cavity, nitroglycerine may be exploded when the potential energy of the fall hammer is as low as 20 g.cm. If the liquid explosive is initially distributed on the anvil in the form of droplets or as parallel strips which coalesce during the impact and entrap gas, the sensitivity is also very high. Even when a continuous film of the explosive is struck between flat surfaces small amounts of gas may be entrapped and may initiate the explosion. If precautions are taken to prevent the entrapping of gas, nitroglycerine is comparatively insensitive to initiation by impact between curved or flat surfaces, and impact energies of the order of 10^2 – 10^6 g.cm. may be necessary to cause explosion. The effect of the gas is mainly a physical one (adiabatic heating), but the chemical nature (oxidizing property) of the gas is also important.

INTRODUCTION

The detonation of solid explosives by mechanical impact has been investigated fairly extensively, but there is still some doubt as to the mechanism, or mechanisms, by which the mechanical energy of the blow initiates the explosive reaction. Liquid explosives may also be detonated by a mechanical blow, but the experimental study of this has not been so extensive, and again, the mechanism of initiation is not fully understood. The study of explosives in the liquid state possesses some advantages since the explosive is homogeneous, and factors such as crystal size, crystalline form, and density of packing, which lead to wide variation in the sensitivity of solid explosives, do not occur. Nevertheless, the impact sensitivity results which are given by various workers for a liquid explosive, such as nitroglycerine, do differ widely. A usual method of measuring sensitivity is to place a film of the explosive on a flat anvil and strike it with a flat-faced weight—in some cases a flat-ended cylinder or disk is allowed to rest on the film, or is suspended above it, and struck by the hammer. The sensitivity may then be expressed in terms of the potential energy of the hammer necessary to give detonation. Berthmann (1941) gives the sensitivity of nitroglycerine on steel surfaces as 14,000 g.cm. Will (1906) gives it as 4000 g.cm. Other values quoted in the literature range from higher than 14,000 down to 500 g.cm. Much of this early work has been undertaken to get

some standard test for sensitivity which would serve as a practical guide to safety in manufacture or operation.

An attempt has been made to investigate in more detail the physical conditions which occur during impact and to determine the factors which are responsible for initiation. Since both the physical and chemical changes which occur during the impact are very rapid, it is necessary to use electrical and optical methods which will record transient phenomena lasting for only a few microseconds. Measurements have been made of the period of contact between the impacting surfaces, of the time during collision at which initiation is observed, and of the rate and mode of propagation of the explosion from the point of initiation through the liquid film.

A factor which was found to be of great importance was the shape of the anvil or striker or the geometrical arrangement of the explosive itself (Bowden, Eirich, Ferguson & Yoffe 1943*a*; Bowden, Eirich, Mulcahy, Vines & Yoffe 1943*b*). It has been shown, for example, that if one of the surfaces contains a small pin-hole or cavity, initiation occurs with very gentle impact. Explosions occur with nitroglycerine when the impact energy is as low as 20 g.cm. If the nitroglycerine is distributed as droplets or parallel strips on the anvil and struck with a flat surface, or if a gas bubble is introduced into a continuous film of the liquid, it is again found that a gentle impact will cause detonation. The same effect is observed with other liquid or gelatinous explosives. It will be shown in this paper that the high sensitivity of liquid explosives is due to the entrapping of minute gas spaces in the explosive.

The size of the gas spaces and the quantity of gas necessary to cause the explosion is very small indeed. A single tiny bubble (the volume at atmospheric pressure may be of the order of 10^{-11} c.c.) is sufficient to initiate the explosion. These minute bubbles are heated by adiabatic compression during impact and form a small nucleus of hot gas which initiates the reaction. The explosion then propagates from this hot nucleus. Unless precautions are taken to prevent the inclusion of such small gas bubbles they readily occur under many conditions of laboratory experiment and practical operation.

This paper is concerned primarily with the impact sensitivity of liquid and gelatinous explosives with different geometrical arrangements of the striker or explosive, and with the part played by adiabatic heating of small entrapped gas bubbles in initiating the reaction. The electrical and photographic study of the time at which initiation occurs and of the propagation of the explosion from the point of initiation is described in the second paper. A third paper deals with the initiation of a liquid explosive by friction.

IMPACT TESTS ON THIN CONTINUOUS FILMS OF NITROGLYCERINE

Experiments have been carried out on the sensitivity of thin continuous films of nitroglycerine to impact between flat surfaces. The nitroglycerine was spread as a film about 2 cm. in diameter and about 3×10^{-3} cm. thick on a lapped, grease-free, brass anvil (see figure 1(*a*)). This value was chosen since it was found that when

the film thickness was greater or smaller than 3×10^{-3} cm. the sensitivity decreased appreciably. The film was struck with a flat brass striker 2.5 cm. in diameter held in a pendulum type of fall hammer similar to figure 2 and to those described in

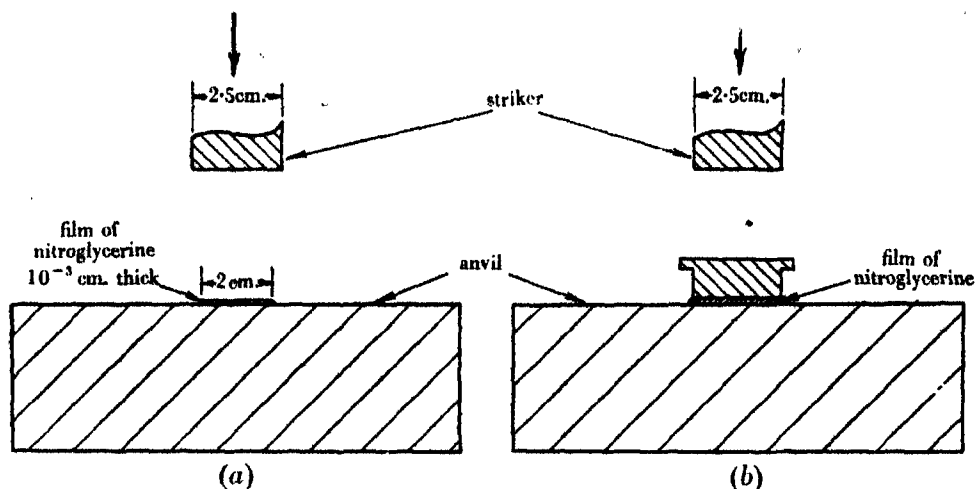


FIGURE 1. (a), nitroglycerine spread as thin film on brass anvil, and struck directly by flat pendulum fall hammer. (b), striker rests on thin film of nitroglycerine on brass anvil and is struck by subsidiary striker held in the hammer. The film may be formed with or without entrapped air spaces.

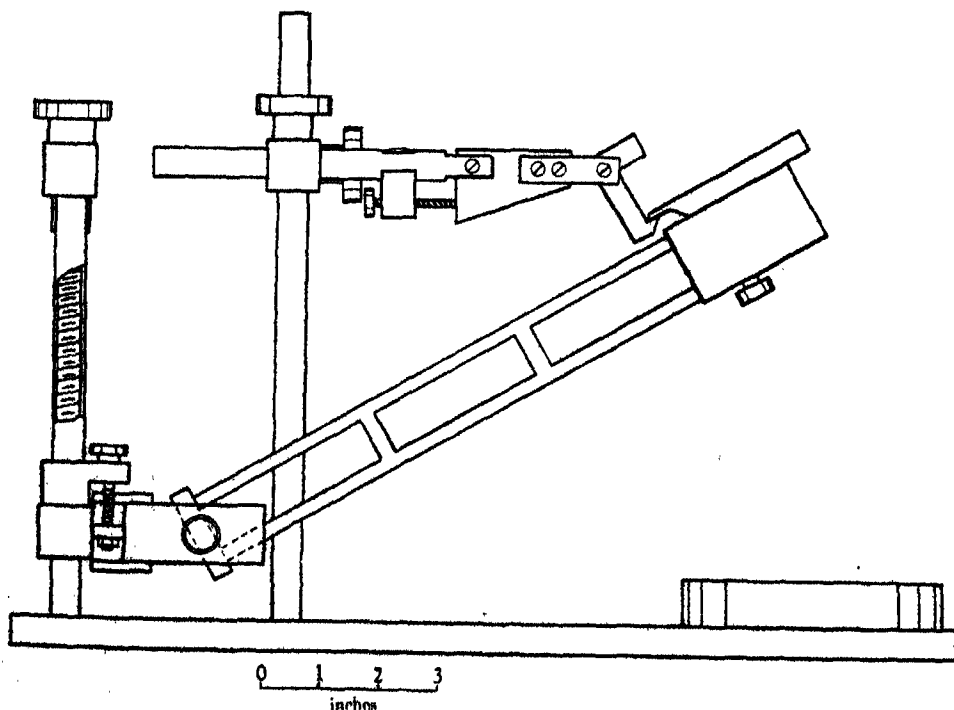


FIGURE 2. Pendulum hammer apparatus.

earlier papers (Bowden *et al.* 1943*a, b*). The striker was levelled on the anvil before each fall. The results obtained for different energies and velocities of impact are given in table 1.

TABLE 1. EXPLOSION EFFICIENCY OF THIN CONTINUOUS FILMS OF NITROGLYCERINE

series	mass of striker (g.)	height of fall (cm.)	energy (g.cm.)	velocity (cm.sec. ⁻¹)	efficiency*
I	1,030	58	60,000	340	30/46 (65 %)
II	290	48	14,000	310	20/35 (57 %)
III	850	21	18,000	200	14/54 (26 %)
IV	380	20	7,600	200	2/20 (10 %)

* Efficiency is defined as (number of explosions)/(number of impacts).

It will be seen from table 1 that the explosion efficiencies obtained in series I and II are similar, although the energies of impact which obtained in series I was four times that in series II. On the other hand, the velocities of approach of the hammer were about the same in both cases. This indicates that the explosion efficiency is determined more by the velocity of approach than by the energy involved. Furthermore, when in series II and III similar energies of impact were used, a high efficiency was observed with the greater height of fall, again demonstrating the importance of velocity of approach.

THE EFFECT OF MINUTE GAS SPACES

The influence of the shape of the impacting surfaces

In view of the variable results obtained with flat impacts the effect of the shape of the striker and of the anvil was studied. Flat strikers of different diameters were used and also curved and conical strikers of varying shapes and sizes. Some of these results will be discussed later, but the most striking observation was that if either the hammer or anvil contained a small pin-hole or cavity less than 1 mm. in diameter the impact sensitivity was enormously increased. The effect was first observed with flat surfaces, but it was more marked with a curved striker containing a small cavity with a raised rim of metal round it. The form of the striker which was used for the majority of the experiments with nitroglycerine is shown in figure 3. It was allowed to fall into a film or lens of liquid explosive on a flat anvil. The striker was first levelled by the adjusting screws on the pendulum hammer. It was judged 'level' when a very light tap on the bare anvil gave a faint but complete circular imprint of the cavity rim. After levelling, the cavity was usually filled by dipping the striker into a drop or film of nitroglycerine on the anvil. The striker was then allowed to fall into the nitroglycerine from the required height. Some typical results with brass surfaces are shown in table 2.

It will be seen that in the presence of a cavity an explosion efficiency of 100 % may be obtained with a weight of 40 g. falling 10 cm. In fact, explosion of nitroglycerine may occur with a 40 g. striker falling 0.5 cm. (i.e. energy 20 g.cm.) corresponding to a velocity of approach of 30 cm.sec.⁻¹. When no cavity is present no explosions are obtained even with a 4 kg. hemispherical striker falling 150 cm., i.e. energy 6×10^5 g.cm. A similar increase in sensitivity was found for other liquid explosives when a cavity was present in the striker. This increase of sensitivity

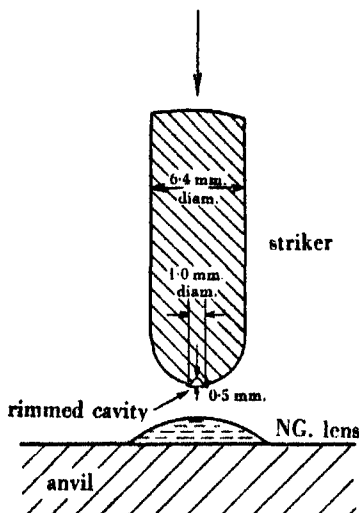


FIGURE 3

TABLE 2. EXPLOSION EFFICIENCY USING A CURVED BRASS STRIKER
WITH A CAVITY AND RAISED RIM ON A BRASS ANVIL

mass of striker (g.)	height of fall (cm.)	energy (g.cm.)	velocity (cm.sec. ⁻¹)	efficiency
40	10	400	140	85/85 (100 %)
40	5	200	99	33/41 (80 %)
40	3	120	78	3/18 (17 %)

is very remarkable: it is evident that liquid explosives may be detonated by the gentlest tap if the anvil or the hammer is of an appropriate shape. To what then is this extreme sensitivity due? One possible explanation is that it is due merely to a concentration of stress at the rim of the cavity. Since contact occurs only at the narrow rim of the cavity all the energy of the impact is concentrated in this small area. Experiments show, however, that this is not the cause. If the cavity is drilled right through the striker so that it is open at the top no initiation occurs, even with the heaviest impacts. Also it is found that when sharp-pointed strikers or curved strikers which give point contact are used no explosion occurs, even with impact energies of the order of 10^6 g.cm. With experiments of this type all

the energy of impact is concentrated into a small area and the stresses are sufficient to cause plastic deformation of brass and steel surfaces. Nevertheless, there is no explosion.

A second possibility is that, with the cavity, initiation is brought about by a tribochemical effect or viscous heating due to the very rapid flow of the liquid as it escapes through the narrowing gap between the rim of the cavity and the anvil. An apparatus for producing high rates of flow was constructed to test this hypothesis. It consisted of a small plunger fitting accurately into a cylindrical hole. When the plunger was struck with a falling weight the liquid explosive was forced out of the bottom of the cylinder and flowed through a narrow gap (thickness about 10^{-3} cm.) between parallel plates. With this apparatus it was possible to get detonation and to reproduce some of the phenomena obtained in the cavity impact. The experiments showed, however, that the flow was of secondary importance. The initiation was really due to the entrapping and compressing of very small gas spaces within the liquid. If precautions were taken to prevent the entrapping of gas no initiation was observed in these flow experiments. This indicated that initiation of the explosion by gentle impact was due to adiabatic compression and heating of entrapped gas bubbles, and a careful examination was made of the cavity experiments to see whether any small gas spaces could be detected there.

Observations under the microscope showed that with the normal experimental procedure a small amount of air is almost invariably entrapped within the cavity. The radius of this bubble may be as small as 5×10^{-3} cm., and it is usually present as a pocket of air in the roof of the cavity and is invisible to the eye. In fact, it may be very difficult to detect with a microscope, unless the cavity is searched with a fine probe and the bubble detached from the wall and brought out into the body of the liquid. It was found that if the air bubble is removed from the cavity the sensitivity to impact disappears. If the bubble is reintroduced the high impact sensitivity is again observed.

Table 3 gives comparative results for the explosion efficiency when air bubbles are deliberately either excluded or included.

TABLE 3

nature of impacting surface	Brinell hardness (kg./mm. ²)	mass of striker (g.)	height of fall (cm.)	energy (g.cm.)	efficiency	
					bubble present	bubble absent
brass	40-80	65	24	1550	15/16 (94 %)	0/15 (0 %)
steel	150-180	35	11	370	60/70 (85 %)	0/80 (0 %)
steel	150-180	200	35	7000	56/59 (95 %)	0/30 (0 %)

It will be seen that a bubble is necessary in order to obtain explosions with energies of impact ranging between 370 and 7000 g.cm.

Experiments were also performed in which the striker was filled with nitroglycerine (with or without a bubble) and lowered into a lens of nitroglycerine on the anvil. In this experiment the nitroglycerine had been previously degassed

under a vacuum pump and the inside of the cavity coated with a layer of shellac in order to reduce the possibility of formation of air bubbles on the metal surface. The striker was held in the lens about 0.75 mm. above the anvil surface by a guide and struck by a weight of 250 g. falling about 30 cm. so that the cavity was completely immersed during the whole of the experiment. The ratios of explosion to impact were:

Bubble in cavity 10/10. No bubble observed in cavity 1/20.

The efficiency of a bubble in causing initiation depends on its size and position in the cavity. Bubbles close to the roof of the cavity are more effective than those suspended near the opening. This, no doubt, is due to the smaller likelihood of a bubble near the roof escaping on impact. For impacts of 195 g. falling 35 cm. (7000 g.cm.), the lower limiting diameter of effective bubbles situated close to the roof of the cavity appears to be about 0.1 mm., i.e. a bubble of diameter less than 0.1 mm. did not initiate explosion at this impact energy. Very large bubbles are also ineffective.

THE PRESSURE AND TEMPERATURE DEVELOPED IN THE BUBBLE DURING IMPACT

The effect of initial gas pressure

The temperature rise in the bubble depends, of course, upon the compression ratio, and if the process is truly adiabatic and the gas ideal, is given by

$$T_2 = T_1 \left(\frac{p_2}{p_1} \right)^{\frac{\gamma-1}{\gamma}}.$$

It is of interest to calculate the temperature rise in cases where the initial pressure p_1 and final pressure p_2 can be estimated, and also to investigate the effect which an alteration of the pressure ratio or of γ has on the incidence of explosion.

A rough experimental estimate of the minimum amount of compression of the bubble in the cavity necessary for initiation may be obtained from a consideration of the results obtained with impacts on anvils of varying hardness (Bowden *et al.* 1943*a*). It has been found that, whereas an initiation is readily effected with cavity brass strikers (Brinell hardness 60–80 kg./mm.²) falling on a lead anvil (hardness 4 kg./mm.²), no explosion has been observed when the anvil is of the softer indium (hardness 1 kg./mm.²), although the indium is plastically deformed by the impact.

The average kinetic flow pressure of indium has been determined and is of the order of 4 kg./mm.² for low-velocity impacts of the order of 50 cm.sec.⁻¹; the average kinetic flow pressure of lead under similar conditions is about 8 kg./mm.². The minimum pressure necessary for initiation which must be developed in the liquid and bubble during the impact must, therefore, lie between 4 and 8 kg./mm.².

If it is assumed that the pressure of the liquid inside the cavity does not exceed the kinetic flow pressure of the metal, an approximate estimate of the pressure may

be made. The maximum temperature rise for an adiabatic compression from atmospheric pressure to 4 kg./mm.² (400 atm.) is of the order of 1300° C. For a final pressure of 8 kg./mm.² (800 atm.) the temperature rise is about 2000° C. We have measured the ignition temperature of nitroglycerine when heated in bulk and find a value of 265–275° C, so that high temperatures may well cause initiation.

Since the initial mass of the gas bubble is about 10^{-9} g., the actual quantity of heat developed in the gas bubble for a temperature rise of 2000° C is about 4×10^{-7} cal.

*Cavity experiments in low-pressure atmospheres of air,
nitrogen, ether and carbon tetrachloride*

A series of experiments was carried out in a vacuum chamber which contained the pendulum fall hammer. This chamber could be evacuated down to moderately low pressures (10^{-2} cm. Hg) or filled with appropriate gases. The arm of the fall hammer could be released from an electromagnet and the experiment observed through a glass window.

In the cavity experiments and in many of the other impact experiments the initial volume of the small entrapped gas bubble is fixed by geometrical consideration. If, before the experiment, the initial gas pressure p_1 is lowered, the ratio of p_2/p_1 will be increased and the temperature rise may be greater. On the other hand, the mass of the entrapped gas and the amount of heat developed will be reduced. In the limit if the initial pressure is very low the mass of entrapped gas and the quantity of heat developed may be so small that it may no longer be capable of initiating the explosion.

The results of carrying out cavity impacts in air at different pressures are shown in table 4. Before use the nitroglycerine was degassed under the vacuum of a hyvac pump. Before each experiment the chamber (containing the nitroglycerine) was evacuated to less than 0.1 cm. and the air then admitted to the required pressure.

TABLE 4. EXPLOSION EFFICIENCY WITH CAVITY IMPACT
IN AIR AT DIFFERENT PRESSURES

Striker 205 g. falling 24 cm. initial air pressure p_1	Energy 4900 g.cm. explosion efficiency
76.0 cm. Hg, 1.00 atm.	31/32 (97 %)
6.4 cm. Hg, 0.08 atm.	40/51 (78 %)
2.0 cm. Hg, 0.03 atm.	5/23 (22 %)
1.0 cm. Hg, 0.01 atm.	0/5 (~0 %)
0.2 cm. Hg, 0.003 atm.	0/10 (~0 %)

It will be seen that the gas bubble within the cavity may be still effective when the initial pressure p_1 is 2.0 cm. Hg. Again, the initial diameter of the bubble could be about 10^{-2} cm. so that its mass could be about 10^{-11} g. Assuming that the final pressure p_2 is about 600 atm., the theoretical maximum temperature rise would be

several thousand degrees. The actual quantity of heat developed within the bubble would, however, have fallen to about 10^{-7} cal. Apparently this small quantity of heat is still capable of initiating the explosion.

If the initial pressure is reduced to a low value ($p_1 = 10^{-3}$ atm.) explosion does not occur, presumably because the amount of entrapped gas and the quantity of heat developed is too small.

The experiments were repeated in atmospheres of ether and carbon tetrachloride vapours. For ether $\gamma = 1.08$ and for carbon tetrachloride $\gamma = 1.13$, so that for both these gases the adiabatic temperature rise of any entrapped bubbles would be greatly reduced. The results are given in table 5.

TABLE 5. CAVITY IMPACT IN GASES OF DIFFERENT γ

Striker 205 g. falling 24 cm. Energy of impact 2900 g.cm.

initial pressure of gas p_1 (cm. Hg)	atmosphere	explosion efficiency
6.4	air $\gamma = 1.4$	40/51 (78 %)
6.4	nitrogen $\gamma = 1.4$	5/5 (~ 100 %)
6.4	ether $\gamma = 1.08$	0/8 (~ 0 %)
6.4	carbon tetrachloride $\gamma = 1.13$	0/8 (~ 0 %)

It will be seen that, under these conditions of impact, the explosion efficiency is reduced from 78 % to zero by the substitution of ether or carbon tetrachloride. The amount of ether which dissolves in nitroglycerine under 6.4 cm. Hg is almost certainly less than the Raoult's law value, i.e. 5 % (wt.), and this consequently would have little desensitizing effect on the nitroglycerine (see later). The solubility of carbon tetrachloride in nitroglycerine would be expected to be even smaller than that of ether, since (as we have observed) nitroglycerine and carbon tetrachloride show only partial miscibility. For ether the theoretical temperature rise would be one-thirteenth and for carbon tetrachloride one-seventh that of air. This result, that gases of low γ reduce the sensitivity, is quite general and supports the view that it is the adiabatic heating of the entrapped gases which is responsible for the initiation.

THE INFLUENCE OF THE DISTRIBUTION OF THE EXPLOSIVE

An increase in the sensitivity of liquid explosives to flat impact may also be obtained by a suitable arrangement of the explosive on the anvil (Bowden *et al.* 1943*b*). If the explosive is distributed as a number of small droplets upon a flat surface or if bubbles of gas are introduced into a continuous thin film of the liquid, explosion may be obtained at relatively low energies of impact.

The size, number and distribution of the droplets are important. A single droplet or a few large drops will not explode even under high energies of impact. It is also found that very small droplets which are spaced widely apart are ineffective. The most effective arrangement consists of a number of small droplets (about 1 mm).

placed closely together. Experiments with a transparent striker show that with this arrangement small air pockets are normally entrapped between the impacting surfaces. Some typical results are given in table 6.

TABLE 6. THE SENSITIVITY OF NITROGLYCERINE TO FLAT IMPACT WHEN DISTRIBUTED AS A NUMBER OF DROPLETS OR WHEN AIR BUBBLES ABOUT 0.5 CM. DIAMETER ARE INTRODUCED INTO A THIN FILM

mass of striker (g.)	height of fall (cm.)	energy (g.cm.)	efficiency		
			droplet distribution	air bubbles in continuous film	continuous thin film
380	20	7600	—	7/7 (100 %)	2/20 (10 %)
180	20	3600	—	13/13 (100 %)	0/7 (~ 0 %)
180	15	2700	50/50 (100 %)	—	—
180	10	1800	— (100 %)	13/21 (62 %)	—
40	15	600	3/25 (12 %)	17/18 (94 %)	0/20 (0 %)

Explosions are readily obtainable with 40 g. falling 15 cm., while for continuous thin films the efficiency at this energy of impact is zero. Again it is suggested that the initiation of the explosion is due to the adiabatic compression of the gas spaces enclosed within the liquid film at the moment of impact.

The fact that a single drop (or a single strip) of explosive causes no increase in sensitivity provides strong support for the view that initiation is not due simply to liquid flow. Unless the geometrical arrangement is such that gas is entrapped during the impact, initiation does not occur at these impact energies.

Further work has shown that even if the amount of gas entrapped is exceedingly small it may still be sufficient to cause explosion. For example, a single bubble of air, having a diameter of 1 mm., in a very thin film of nitroglycerine, is capable of causing an explosion at an impact energy of 600 g.cm. (40 g. falling 15 cm.).

The effect of the gas spaces depends not only on the adiabatic compression; the chemical nature of the gas is also important (see later). For instance, the sensitivity of the explosive with oxidizing gases such as air, oxygen, nitrous oxide, is much greater than when inert gases such as nitrogen and carbon dioxide are used. This increase in sensitivity has been assumed to be the result of an initial reaction with the excess oxygen (possibly in the vapour phase) in which a large amount of energy is released and an explosion nucleus thereby created.

Further evidence that extremely small quantities of entrapped gas, either inert or oxidizing, are capable of producing explosion of liquid nitroglycerine on impact has been obtained from experiments with parallel strips of nitroglycerine.

The explosion of parallel strips of nitroglycerine

An increase in sensitivity even greater than that observed when air bubbles are introduced previous to the impact has been obtained when the nitroglycerine is simply distributed as two thin parallel strips on a flat anvil as shown in figure 4.

The strips were struck with a very flat brass striker about 1.5 cm. in diameter. Clean freshly lapped surfaces were used for each experiment. The results obtained under these conditions are shown in table 7 and compared with corresponding experiments in which the nitroglycerine was spread on the anvil as a continuous film.

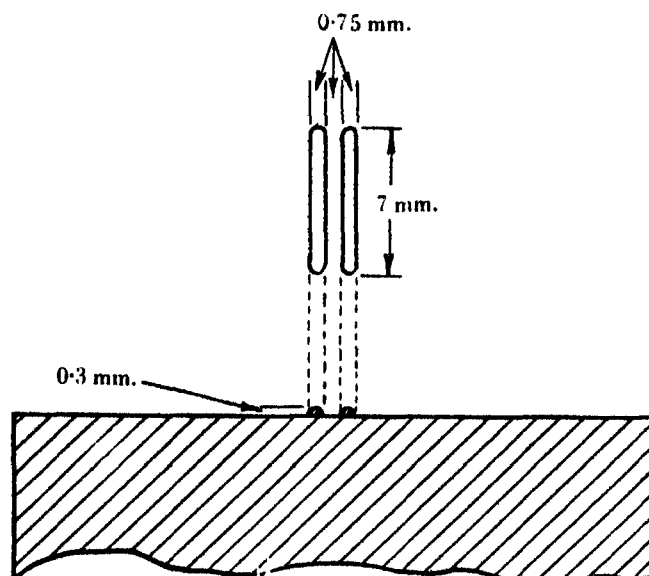


FIGURE 4. Nitroglycerine distributed as strips.

TABLE 7

mass of striker (g.)	height of fall (cm.)	energy (g.cm.)	efficiency	
			strips	continuous thin film
380	20	7600	—	2/20 (10 %)
230	20	4600	22/25 (90 %)	—
55	18	1000	28/39 (70 %)	—
42	15	630	6/7 (80 %)	0/20 (0 %)
42	10	420	4/10 (40 %)	—
41	8	330	4/10 (40 %)	—
41	6	250	2/25 (10 %)	—

Thus, as in the case of the experiments with droplets and bubbles, there is a marked increase in the sensitivity to flat-surface impact when the nitroglycerine is distributed as strips. With an impact energy of 7600 g.cm. the efficiency for continuous thin films is only 10 %, while with strips the efficiency is 90 % at 4600 g.cm. The lower limit of sensitivity with the strips appears to be 250 g.cm., which is not very different from that observed with cavity impacts (20 g.cm.).

The presence of two or more strips is necessary to obtain the increase in sensitivity. When a single strip is used no increase is observed. This supports the view that initiation is not due to liquid flow but to the adiabatic compression of gas bubbles entrapped in the nitroglycerine during the impact.

In these experiments several extremely small bubbles may be entrapped, the volume of each being about 10^{-4} c.c. and its mass 10^{-7} g. At medium energies of impact (e.g. 400 g. falling 15 cm.) for continuous thin films between flat surfaces, Cherry (1945) has estimated the maximum pressure developed in the liquid during impact to be of the order of 1000 atm., while at very low energies of impact (40 g. falling 10 cm.) Eirich & Tabor (1945) have estimated the pressure to be 20 atm. The maximum pressure is developed at the centre of the striker. Assuming the pressures to be of the same order when small bubbles are present and that the gas is ideal even at 1000 atm., the volume of each bubble would decrease at these pressures to 10^{-7} and 3×10^{-6} c.c. respectively. The corresponding calculated temperature flashes due to adiabatic compression are 2000 and 400° C, and the quantities of heat developed are of the order of 3×10^{-5} and 7×10^{-6} cal. respectively. The calculation by Eirich & Tabor of the maximum pressure developed in the liquid indicates that the high temperatures, even for the low-energy impacts, may be reached during the first stages of the collision. This is in accord with the observation that explosions of nitroglycerine distributed as strips may occur at the initial phase of the collision, long before the liquid film has been squeezed out appreciably (see next paper).

Impact on parallel strips in atmospheres of air, nitrogen, hydrogen and ether

A series of experiments similar to those described earlier for cavity impacts was carried out with a flat striker on parallel strips of nitroglycerine. The vacuum box was evacuated to 10^{-1} cm. Hg and then filled with the appropriate gas. For the experiments with nitrogen and hydrogen the chamber was first flushed out three times with the gas and then filled to the required pressure. In the ether experiments the vapour was introduced by allowing a volume of the liquid to evaporate. It should be pointed out that in these particular experiments a small amount of air, about 10^{-1} cm. Hg, remained in the vessel. The results obtained are shown in table 8.

TABLE 8. COMPARISON OF EXPLOSION EFFICIENCY OBTAINED WITH STRIPS IN LOW-PRESSURE ATMOSPHERES OF AIR, NITROGEN, HYDROGEN AND ETHER

initial pressure of gas (cm. Hg)	Weight 230 g. Height 20 cm.			
	efficiency			
	air	nitrogen	hydrogen	ether
76	36/40 (90%)	14/16 (88%)	4/17 (24%)	—
10.0	27/32 (85%)	5/10 (50%)	12/27 (44%)	0/10 (0%)
3.5	22/26 (85%)	8/16 (50%)	—	3/20 (15%)

Again it will be seen that the explosion efficiency in ether ($\gamma = 1.08$) is less than in nitrogen, hydrogen and air at the same pressure. Assuming that the final pressure developed in the liquid (and the entrapped gas spaces) during the impacts is about 1000 atm., the maximum theoretical temperature resulting from the compression of ether at an initial pressure of 10 cm. Hg would be 270°C . When the initial pressure of ether is 3.5 cm. Hg the temperature could be 310°C . The corresponding estimated maximum temperature, resulting from the compression of air, nitrogen or hydrogen from the same initial pressures, could be about 8000 and 5000°C respectively. Since the figure given for the final pressure is uncertain, these simplified calculations should be considered as a comparative indication only.

It is also evident from table 8 that the explosion efficiency in a nitrogen and hydrogen atmosphere at a pressure of 10 cm. Hg is lower than in air at the same pressure. In view of the similar value of γ for nitrogen, hydrogen and air it is clear that the adiabatic compression of the gas represents only one of the factors operating in the initiation mechanism. The results suggest that, in addition to the physical effect of the gas in causing adiabatic heating, its chemical nature (oxidizing properties) is important. Similar behaviour has been found in experiments with nitroglycerine films containing gas bubbles.

Some additional experiments were carried out in air at a still lower pressure (10^{-2} cm. Hg), and the results are given in table 9.

TABLE 9. EXPLOSION EFFICIENCY OBTAINED WITH STRIPS IN AIR
AT LOW PRESSURE AND AT ATMOSPHERIC PRESSURE

mass of striker (g.)	height of fall (cm.)	energy of impact (g.cm.)	efficiency in air	
			at pressure of 10^{-2} cm. Hg	at atmospheric pressure
230	20	4600	12/14 (86 %)	36/40 (90 %)
45	14	630	5/5	3/3

It is clear that the explosion efficiency is still high when the air pressure is lowered to 10^{-2} cm. Hg. At this pressure the amount of air entrapped, and consequently the quantity of heat developed, must be extremely small. Simple calculations show that the mass of gas in each bubble entrapped at 10^{-2} cm. Hg is about 2×10^{-11} g. The calculated temperature rise obtained from adiabatic compression from this initial low pressure is some twenty times greater than that from compression at atmospheric pressure. It seems, therefore, that it is the actual temperature rise rather than the quantity of heat developed (or perhaps a combination of these two factors) which is important in initiating the explosion at the low energy of impact (630 g.cm.). The quantity of heat developed in the bubble may be calculated and is about 10^{-7} cal. The reason for the different behaviour of cavity impacts in this respect is not clear, but may be associated with the smaller amount of air which is entrapped by the cavity.

Droplet experiments in low-pressure atmospheres of oxygen, nitrogen and air

A series of experiments similar to those described in the previous section was carried out with nitroglycerine distributed as droplets in low-pressure atmospheres of oxygen, nitrogen and air. The droplets were spread on the anvil which was previously coated with a very thin film of grease. Such a film does not affect the sensitivity to impact in the case of liquid explosives. It has been shown in a previous paper (Bowden *et al.* 1943*b*) that for the experiments with droplets the explosion efficiency is determined mainly by the velocity of approach of the striker and not primarily by the energy of the blow. It was found that by taking velocity of approach as criterion, the order of explosion efficiency in the different gaseous atmospheres is:

low-pressure oxygen (10 cm. Hg) > air at atmospheric pressure
 > low-pressure atmosphere of air (10^{-3} cm. Hg) > nitrogen at atmospheric pressure.

These results support those obtained from the experiments with nitroglycerine distributed as strips, in that they show the presence of an effect due to the chemical nature of the atmosphere; explosions are more readily obtained in air at atmospheric pressure than in nitrogen at the same pressure. Furthermore, the results show that explosions may be obtained when the initial pressure of the gas is as low as 10^{-3} cm. Hg.

THE ACCIDENTAL TRAPPING OF GAS SPACES DURING FLAT IMPACT

Figures for the explosion efficiency of thin continuous films of nitroglycerine to flat-surface impact have already been given in an earlier section. When transparent strikers or anvils were used it became evident that under the normal conditions of flat-impact experiments, air bubbles are readily trapped in the nitroglycerine film during the impact of the surfaces. As pointed out above, the presence of entrapped gas has a profound effect on the sensitivity. The question therefore arises as to how far the impact sensitivity of nitroglycerine reported in the literature and in table 1 is influenced by this factor. It might explain in part the wide discrepancies between the values obtained by different workers. The experiments with glass strikers and anvils showed that even at low velocities of approach of the striker it was almost impossible to exclude air spaces. The results described above suggest that the adiabatic compression of these gas spaces may be responsible for initiation. In order that this might be tested, further experiments were carried out under conditions in which air bubbles were (a) intentionally entrapped and (b) removed.

In these experiments the striker did not fall directly on the nitroglycerine film. In the (a) series, a thin film of nitroglycerine was placed on the anvil and a brass disk allowed to rest on the top of it. The fall hammer then struck the top disk (see figure 1*b*). Parallel experiments with a glass disk showed that almost invariably

some air was entrapped between the disk and the anvil previous to impact. In the (b) experiments a thick lens of nitroglycerine was placed on the carefully cleaned anvil. The clean brass disk was also wetted with a thick film of the explosive and then lowered slowly and obliquely on to the anvil. The excess nitroglycerine was sucked away with a syringe. Again parallel experiments with a glass disk showed that air bubbles were *usually* eliminated by this technique.

With a fall hammer of 1 kg. falling some 60 cm. (energy of impact 60,000 g.cm.) the results for the explosion efficiencies were as follows:

- (a) gas spaces intentionally entrapped 10/10 (100 %),
- (b) gas spaces probably absent 2/35 (6 %).

This figure of 60,000 g.cm. may be taken as an upper limit only, since it was found that with this particular hammer a considerable amount of the energy was absorbed in plastic deformation of the rear end of the brass striker.

The results show that the presence of a gas space greatly increases the probability of explosion.

There is evidence that if the impact energy is very great the nitroglycerine may be detonated by impact in the absence of gas spaces. A layer of nitroglycerine was enclosed between a brass disk and a glass slide (arrangement similar to figure 1b) and the film closely scrutinized to ensure that no bubble was entrapped. The glass slide was 'stuck' to a brass anvil with a thin layer of nujol and the disk subjected to the impact of 5 kg. falling 120 cm.—energy of impact about 10^6 g.cm. Regular explosions were obtained. In this case, initiation may be due to shock or to a temperature rise within the rapidly flowing liquid.

Recently, Cherry (1945) has made a mathematical analysis of the liquid flow between two approaching surfaces. He has shown that, if allowances are made for the elastic deformation of the approaching surfaces, the temperature rise in the liquid film under moderate conditions of impact is only a few degrees Centigrade. Eirich & Tabor (1945) arrive at the same conclusion. According to Cherry's calculations there is very little actual 'shock' on impact, since before contact the surfaces suffer considerable elastic deformation. This deformation absorbs the major part of the energy of impact so that only a small proportion is actually absorbed by the explosive liquid; consequently the rise in temperature within the film at the lower energies of impact (up to 10^4 g.cm.) is in most cases no more than 50°C . It is only at very high energies of impact (about 10^6 g.cm.) that the theoretical temperature obtained during the shearing of the liquid as it escapes between the surfaces reaches several hundred degrees, near the periphery of the striker. It is therefore improbable at low or moderate energies of impact that the initiation is due to a viscous heating of the rapidly flowing liquid. This can only be true at very high energies of impact, e.g. 10^5 – 10^6 g.cm.

In view of the effect of air bubbles entrapped in thin films, it is interesting to note that in previous impact experiments (Bowden *et al.* 1943*a*), where the striker fell upon an open film of nitroglycerine, it was found that strikers of large area

are more efficient than small ones. With brass strikers 0.6 cm. in diameter it is very difficult to initiate the explosion at all, even with very high energies of impact. If the diameter is increased to 2.5 cm., both the sensitivity and the probability of explosion are very greatly increased. With the striker of greater diameter, the probability of entrapping air is obviously greater.

Also it is found that very thick films of the liquid are less sensitive than thin ones. In thick films the bubbles would be more mobile and could more readily escape before the pressure rose to a high value. If the film on the anvil is too thin, the sealing and compression are inefficient and again the probability of detonation is lower.

THE DETONATION OF LIQUID EXPLOSIVE OTHER THAN NITROGLYCERINE

Previous work (Bowden *et al.* 1943*b*) has shown that provided an explosive is in the liquid state, either molten or dissolved in solvents, its behaviour is similar to nitroglycerine, both as regards its high sensitivity to cavity impacts and its propagation mechanism.

Nitroglycol, as might be expected, is extremely sensitive to the impact of a cavity striker, an explosion efficiency of 100 % being obtained at an energy of impact of 300 g.cm., i.e. 30 g. falling 10 cm. Gas-space experiments with nitroglycol (bubbles, strips and droplets), show that, in these circumstances, its behaviour is hardly distinguishable from that of nitroglycerine. Similarly, T.N.T., picric acid, trinitroanisole, tetryl, P.E.T.N. and R.D.X., at temperatures just above the melting-point, are sensitive to cavity impacts of an energy of about 300 g.cm.—30 g. falling 10 cm. At these low energies, however, the propagation of the explosion is very limited. T.N.T. may also be decomposed by cavity impact when it is dissolved in various solvents. For example, using nitrobenzene as a solvent at 50° C, a 45M % T.N.T. solution may be obtained, and decomposition occurs with an impact energy of 1000 g.cm. Using a 35M % T.N.T.-dioxan solution at room temperature, decomposition was obtained at an impact energy of about 5000 g.cm. The effect of solvents such as toluene, benzene, ethyl alcohol and nitrobenzene on the sensitivity of nitroglycerine to cavity impact was also studied, and it was shown that the explosion efficiency was still 100 % with a 62M % nitroglycerine-toluene mixture at an energy of impact of 500 g.cm. Blasting gelatine (9 % nitrocellulose), when spread as thin films on a metal anvil and subjected to a cavity impact, gave an explosion efficiency of 90 % with a 40 g. striker falling 15 cm. (energy 600 g.cm.), and explosions were obtained with impact energies as low as 70 g.cm. A similar result was also obtained with dynamite. Further work has been done with five other liquid explosives—methyl nitrate, ethyl nitrate, tetranitromethane, nitromethane and diglycol dinitrate.

Methyl nitrate and ethyl nitrate. The sensitivity of methyl nitrate to impact in the presence of gas spaces is of the same order as that observed for nitroglycerine. Using a 40 g. cavity striker, explosions could be obtained from a height of fall of

1 cm. (energy 40 g.cm.). When flat surfaces were used and a small air bubble introduced into a thin film of the explosive, it was possible to obtain explosions with a 60 g. striker falling 5 cm. (energy 300 g.cm.). Most of the explosive confined between the impacting surfaces was consumed. This high sensitivity may be contrasted with the behaviour of methyl nitrate when spread as a thin film and subjected to impact between flat steel surfaces when no gas spaces are present. Under these conditions no explosions were obtained with 1.8 kg. falling 100 cm. (energy 1.8×10^5 g.cm.).

Ethyl nitrate, on the other hand, was found to be much less sensitive to impact when cavity strikers were used. The lower limit for the sensitivity appears to be about 800 g.cm. Even at high energies of impact the explosion did not propagate and the reaction was confined to the cavity.

Tetranitromethane was interesting for, although it behaved in a fashion fundamentally similar to that of nitroglycerine, there were marked differences. It was found that pure tetranitromethane was remarkably insensitive to flat-surface impact, cavity impact or spark. It is well known that mixtures of this substance with hydrocarbons are explosive, and on mixing it with a small quantity of toluene it was in fact readily initiated by a spark and gave a loud and powerful explosion (cf. Sidgwick 1937). Furthermore, the mixture was found to be sensitive to cavity impacts of the order of 900 g.cm. This reactivity is presumably due to the combustion of the toluene, the tetranitromethane acting as an oxidizing agent. If the volume composition of the mixture is less than about 12 % toluene, no propagation of explosion initiated by sparking is observed. Above 12 %, explosions are observed which are remarkable for their brisance, and which appear to develop, at least at low toluene percentages, in a manner similar to that of nitroglycerine. When the concentration of toluene is greater than 60 %, the passage of a spark no longer produces an explosion. Explosion by impact in the absence of gas bubbles is difficult to obtain (cf., however, the work of Stettbacher (1930), who reports an extremely high sensitivity to flat impact). Flat-surface impacts on thin films of mixture failed to produce an explosion, even at energies of the order of 10^4 g.cm. As mentioned before, however, when the striker contained a cavity as in figure 3, explosion occurred at impact energies of 900 g.cm., but the efficiency was low.

Nitromethane. It was possible to obtain explosions with a 40 g. cavity striker falling 10 cm. (energy 400 g.cm.). When small air bubbles were introduced into a thin film of nitromethane, and the explosive then subjected to flat impact, explosions were observed with a 60 g. striker falling 20 cm. (energy 1200 g.cm.). The explosions obtained were weak and did not propagate to any considerable extent.

Diglycol dinitrate. Using cavity strikers, an explosion efficiency of 10 % was observed with a 40 g. striker falling 10 cm. (energy 400 g.cm.). When small air bubbles were introduced into thin films of the explosive and subjected to flat impact, explosion occurred with a 50 g. striker falling 12 cm. (energy 600 g.cm.).

It is apparent that all the liquid explosives which have so far been investigated are very sensitive to impact when small gas spaces are present in the explosive.

It is possible that the high sensitivity of methyl nitrate when compared with ethyl nitrate is due to its high vapour pressure, and this suggests that the initiation of the explosion may occur in the vapour phase.

These particular experiments have been confined to liquid and plastic explosives, but it is possible that the inclusion and adiabatic compression of small gas spaces may play a larger part in the initiation of solid explosives than is generally recognized.

DISCUSSION

The experiments show that impact sensitivity of nitroglycerine (and other explosives in the liquid state) is increased enormously if tiny gas spaces are present in the liquid or are trapped in it during the impact. This effect is illustrated diagrammatically in figure 5, where the energy required to initiate the explosion is plotted against the explosion efficiency. The amount of air which is capable of bringing about the increase in sensitivity is minute. In the experiments where the explosive is distributed as strips, the entrapped air spaces may have a volume of about

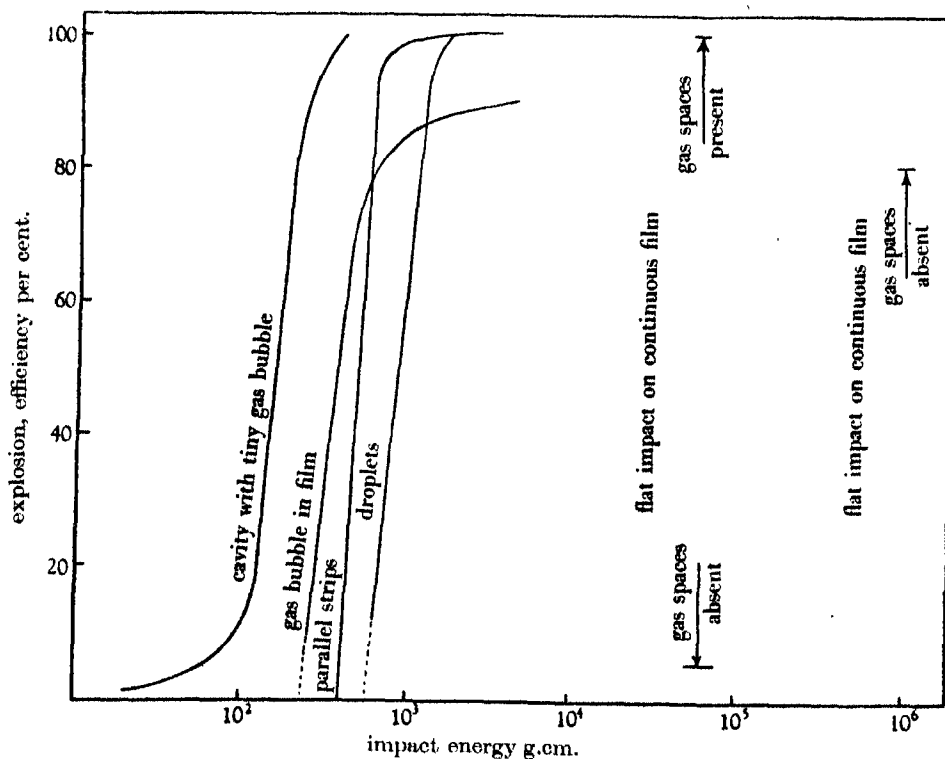


FIGURE 5

10^{-4} c.c. (mass 10^{-7} g.) before compression, and these will cause explosion with a low impact energy (250 g.cm.). The undiminished sensitivity of the strips when struck in a low-pressure atmosphere shows that the mass of air required to sensitize the explosive may be as little as 10^{-10} g. and the volume when compressed about 10^{-11} c.c.

The size of the bubbles at the first instant of their formation is roughly determined by the geometry of the arrangement of the strips, and may be assumed to be independent of the initial pressure of the air.

There is little doubt that the sensitivity is associated with the adiabatic compression and heating of the air bubble under the impact. However, the high sensitivity at low pressure of air indicates that the quantity of heat developed in the air bubble is a relatively unimportant factor compared with the actual temperature rise in the range investigated. For example, the ratio of the quantities of heat generated by equal compression of similar bubbles from atmospheric pressure and from 10^{-2} cm. Hg may be crudely calculated by multiplying the estimated rises in temperature by the mass of air in each case. Reduction of the pressure to 10^{-2} cm. causes a reduction in mass of the bubble by 1/7500 and results in an increase of the temperature rise by 20. The quantity of heat generated under the low-pressure condition is therefore only about 1/370 of that obtained by equal compression of a bubble from atmospheric pressure. It is possible, of course, that the quantity of heat generated under the low-pressure conditions, though much less than that obtained at atmospheric pressure, is still above the threshold quantity necessary for initiation under these conditions. The fact that in the cavity experiments a decline in explosion efficiency with diminishing pressure was observed is evidence of this. The lower sensitivity in the ether atmosphere than that in nitrogen at the same initial pressure (ratio of temperature rises 1:20), for both strip and cavity experiments, indicates strongly that a fairly high temperature must be reached in the bubble for initiation to be effected.

Again, though it is clear that it is the high temperature reached during the compression of the included air spaces which brings about initiation, it is also evident that the high increase in sensitivity is also due in part to the chemical nature of the gas, viz. to the presence of oxygen. This result emerges from the fact that, if included spaces of nitrogen are substituted for air spaces in the liquid, the explosion efficiency falls off (although the sensitivity is still enormously greater than that obtained in the absence of gas spaces). The importance of the chemical (oxidizing) nature of the included gas has been shown by the results of experiments previously reported (Bowden *et al.* 1943*b*). It was found that the introduction of gas bubbles into the liquid previous to impact greatly increases the sensitivity, but that the explosion efficiencies obtained when nitrogen or carbon dioxide is used are considerably less than when air, oxygen or nitrous oxide is introduced under the same conditions of impact. Further evidence that quite high temperatures are reached inside the bubble during compression is obtained from the electrical and photographic study of the initiation and propagation of the explosion

which will be described in the next paper. It will be shown that initiation of the explosion, whether due to compression of a bubble within a cavity or within a thin film of the liquid, takes place within 50×10^{-6} sec. from the beginning of the exertion of the pressure. This very short time-lag is an indication that high-temperature flashes are reached inside the bubble. A similar conclusion has been arrived at from the theoretical calculations of the pressure produced in the liquid during impact. It is possible that initiation of the explosion occurs in the vapour phase inside the bubble.

It would, of course, be interesting to calculate the time lag between the application of the high-temperature flash to the onset of decomposition. Unfortunately, no figures are available in the literature for temperatures of 400°C and greater. Roginsky (1932) did show that the time lag below 200°C could be expressed by $\tau = 1.5 \times 10^{-10} e^{(25,700/RT)}$. This relation has not been tested above 200°C . At a temperature of 900°C , which corresponds approximately to the calculated temperature rise of the air bubble in the cavity experiments, τ , according to the above equation, would be 10^{-5} sec. Our photographic and electrical study indicates a time lag from 0 to 50×10^{-6} sec.

The readiness with which small air spaces are entrapped in a continuous film of nitroglycerine under ordinary conditions of impact, and the large decrease in sensitivity observed when special precautions are taken to reduce this possibility indicate that the high sensitivity (a few thousand g.cm.) recorded for nitroglycerine in the literature may, in fact, be due to this cause. Furthermore, it appears from the few experiments in which the possibility of the trapping of air spaces during impact has been avoided, that it is difficult to obtain explosions with an impact energy of about 6×10^4 g.cm. At these large energies of impact, Cherry (1945) has shown that the temperature rise due to the viscous flow of nitroglycerine may amount to several hundred degrees, and it is possible that initiation of the explosion at these high energies is due to this effect. It would seem, therefore, that if the inclusion of these gas spaces is avoided, nitroglycerine is an explosive which is comparatively insensitive to detonation by impact.

The fact that a wide range of explosive substances is rendered sensitive to gentle impact if small bubbles are present is of some practical interest. Since the bubbles can be so small that they are difficult to see even with a microscope they readily escape detection, and, under suitable conditions, they may constitute a serious hazard.

We wish to express our thanks to Sir David Rivett, F.R.S., and to the Executive of the Council for Scientific and Industrial Research for their stimulating interest and encouragement in the work described in these papers, to the University of Melbourne and to Professor E. J. Hartung for providing facilities and accommodation, and to Imperial Chemical Industries of Australia and New Zealand Ltd and the Munitions Supply Laboratories for providing explosives. We are greatly indebted to Mr A. E. Ferguson, B.E.E., for the design of much of the electrical apparatus, and

to Mr J. S. Courtney-Pratt, B.E., and Mr T. V. Krok, B.E., for assistance with the design of the fall hammers and camera. Our thanks are also due to Mr J. R. Richards, M.Sc. and Mr N. N. Greenwood, B.Sc. for assistance with some of the experimental work.

REFERENCES

- Berthmann, A. 1941 *Chem. Abstr.* **35**, 625.
Bowden, F. P., Eirich, F., Ferguson, A. E. & Yoffe, A. 1943*a* *Bull. Coun. Sci. Industr. Res., Aust.*, no. 167.
Bowden, F. P., Eirich, F., Mulcahy, M. F. R., Vines, R. G. & Yoffe, A. 1943*b* *Bull. Coun. Sci. Industr. Res., Aust.*, no. 173.
Cherry, T. 1945 *Rep. Coun. Sci. Industr. Res., Aust.*, **A**, **116**, no. 8.
Eirich, F. & Tabor, D. 1945 *Rep. Coun. Sci. Industr. Res., Aust.*, **A**, **121**, no. 9.
Roginsky, S. Z. 1932 *Phys. Z. Sowjet.* **1**, 640.
Sidgwick, N. V. 1937 *The organic chemistry of nitrogen*. Oxford University Press.
Stettbacher, A. 1930 *Z. ges. Schiess- u. Sprengstoffw.* **25**, 439.
Will, W. 1906 *Z. ges. Schiess- u. Sprengstoffw.* **1**, 209.
-

The period of impact, the time of initiation and the rate of growth of the explosion of nitroglycerine

By F. P. BOWDEN, M. F. R. MULCAHY, R. G. VINES AND A. YOFFE, *Tribophysics Section, Council for Scientific and Industrial Research, University of Melbourne*

(Communicated by Sir David Rivett, F.R.S. and Sir Robert Robertson, F.R.S.—
Received 12 January 1946—Read 9 May 1946)

[Plates 5–9]

Electrical and photographic methods have been used to measure the period of impact between colliding solids and to study the time of initiation and the rate of growth of an explosion produced by the impact. The measurements were made with liquid explosives, where, as the last paper has shown, the initiation is due to the entrapping of small gas bubbles. Under the conditions of these experiments, the period of impact—that is the time from the first moment of contact of the solid surfaces until they separate on the rebound—is about 200 μ sec.

If one of the surfaces contains a cavity, initiation with nitroglycerine is first apparent as a tiny spot of light inside the cavity. This occurs at the first moment of contact or a few microseconds after. This small flame spreads slowly through the cavity with a velocity of about 20 m./sec., and after a short interval of time (about 50 μ sec.) bursts through the cavity wall and explodes the main film. The explosion is propagated through the film with a velocity > 1000 m./sec., and the residual burning inside the cavity is finally extinguished when the surfaces separate.

When the initiation is due to a bubble entrapped in the coalescing liquid, the measurements show that initiation again begins as a small flame which becomes visible at the moment of coalescence or a few microseconds afterwards. Both the coalescence and the initiation can

occur when the solid surfaces are still separated by a film of explosive. The flame is propagated outwards with a circular symmetry at a velocity of about 400 m./sec., and after a short distance it transforms into a higher velocity and is propagated through the thin film of explosive at a velocity > 1000 m./sec.

These results support the conclusions of the last paper that the explosion is initiated at a small gas bubble which has been heated to a high temperature by adiabatic compression. This bubble of hot gas brings about the thermal decomposition of the explosive: the reaction begins as a comparatively gentle burning which, after a few microseconds, passes over into a more violent explosion.

INTRODUCTION

The last paper described experiments on the sensitivity of nitroglycerine and other liquid and gelatinous explosives to impact between metal surfaces. It was shown that the high sensitivity is due to the presence of small gas spaces entrapped in the explosive and heated by adiabatic compression during the impact. These gas bubbles are regularly entrapped when there is a small cavity in the surfaces or when the explosive is distributed as strips or droplets and they are frequently entrapped during the impact of flat surfaces on a continuous film of the explosive.

This paper describes an attempt to determine at what stage of the impact initiation occurs and also to follow the course of the explosive reaction as it grows from the point of initiation at the tiny hot bubble, spreads through the liquid film and develops into a large-scale detonation.

In the case of impact between surfaces containing a cavity, for example, measurements have been made of the time for which the solid surfaces are actually in contact, and the moment during the impact at which initiation is first apparent has been determined. A similar measurement has been made of the rate of coalescence of strips of explosive as they flow together under the impact of flat surfaces and of the time at which initiation occurs. The rate of growth of the initial tiny explosion and its propagation into the surrounding film has also been studied. The cavity itself is very small (about 0.4 mm. radius), and a specially designed moving-film camera was used for investigating the development of the explosion within it. Cathode-ray methods and a high-speed electronic counter were also used for studying the sequence of events.

The first part of the paper deals with cavity impact, and the second with the impact of flat surfaces on strips and droplets of the explosive.

PART I. THE SUCCESSIVE STAGES OF EXPLOSION INITIATED BY CAVITY IMPACT

In the earlier work (Bowden, Eirich, Ferguson & Yoffe 1943*a*; Bowden, Eirich, Mulcahy, Vines & Yoffe 1943*b*), an experimental study of the initiation and propagation of the explosion was made (i) by observation of the blast marks found on hammer and anvil after explosion, (ii) by photographing the light effects manifested about the cavity and within it during the explosion, and (iii) by measurements of the electrical conductivity between hammer and anvil during impact and explosion.

If a cavity striker is allowed to fall into an extended thin film of explosive confined between two flat surfaces, the explosion may be carried over from the cavity and propagated throughout the film. Examination of the blast marks left on the confining surfaces and of the light effects recorded when one of the surfaces is a thin sheet of mica resting on a photographic plate, shows that under these conditions the explosion is propagated in at least two successive stages. Figure 1*a*, plate 5 shows the marks left on the brass surface after such an explosion. The point of initiation is in the region of the imprint of the rim of the cavity which is seen as a small circle in the centre. The explosion has spread out radially from here but has left the surface substantially untouched until it reached a critical radius C , where heavy deformation of the surface begins. This indicates that the process has been transformed suddenly into a second stage of explosion different in nature from that which prevailed over the central unmarked area. The two-stage behaviour is not peculiar to initiation by cavity impact but is general, since it takes place whether the initiation is affected by impact between flat surfaces, by electric spark or by heat.

If one of the confining surfaces is a sheet of mica on a photographic plate, the light effects recorded on the photographic plate correspond closely to the blast patterns which remain on a brass surface. Figure 1*b*, plate 5 is a print from such an experiment and shows the features characteristic of cavity initiation and subsequent propagation into the confined film, namely, the bright illumination in the cavity and, beyond the cavity, the initial region of weak illumination during the first stage of the propagation which is suddenly converted into the second stage of brighter illumination corresponding to the regions where deformation of the surfaces occurs.

As mentioned above, the photographic technique has now been extended to the use of a moving film with the object of ascertaining the sequence of the several separate events recorded by the photographic plate (e.g. in figure 3, plate 6) and of measuring the velocity with which the explosion is propagated.

Experimental

The nitroglycerine was confined as a thin layer between an upper flat brass surface and a sheet of mica (or glass). The mica in this case rested on the upper surface of a rigid rectangular brass block which was constructed so that it was bisected along about 10 cm. of its long axis by a narrow slit. The slit was filled with 'perspex' which was tightly clamped in, lapped flush with the upper brass surface, and polished at both its upper and lower surfaces to allow the optimum transmission of light. In this manner the perspex acted as a transparent medium and at the same time provided rigid support for the explosive. The upper confinement was provided by the weight, 500 g., of the rectangular piece of brass resting on the explosive. The slit was set at right angles to the direction of rotation of a drum carrying a photographic film on which the image of the slit was focused. The

explosion was initiated by allowing a cavity striker to fall through a narrow hole, drilled at one end in the upper confining surface, on to a small reservoir or lens of nitroglycerine resting on the mica sheet. Under these conditions the explosion propagated along the length of the confined film. A section of the apparatus (not drawn to scale) is given in figure 4. The apparatus could also be used with a different upper plate for spark-initiated detonation. The peripheral speed of the camera could be varied from 12.5 to 25 m./sec. The film used was Eastman Super XX (35 mm.).

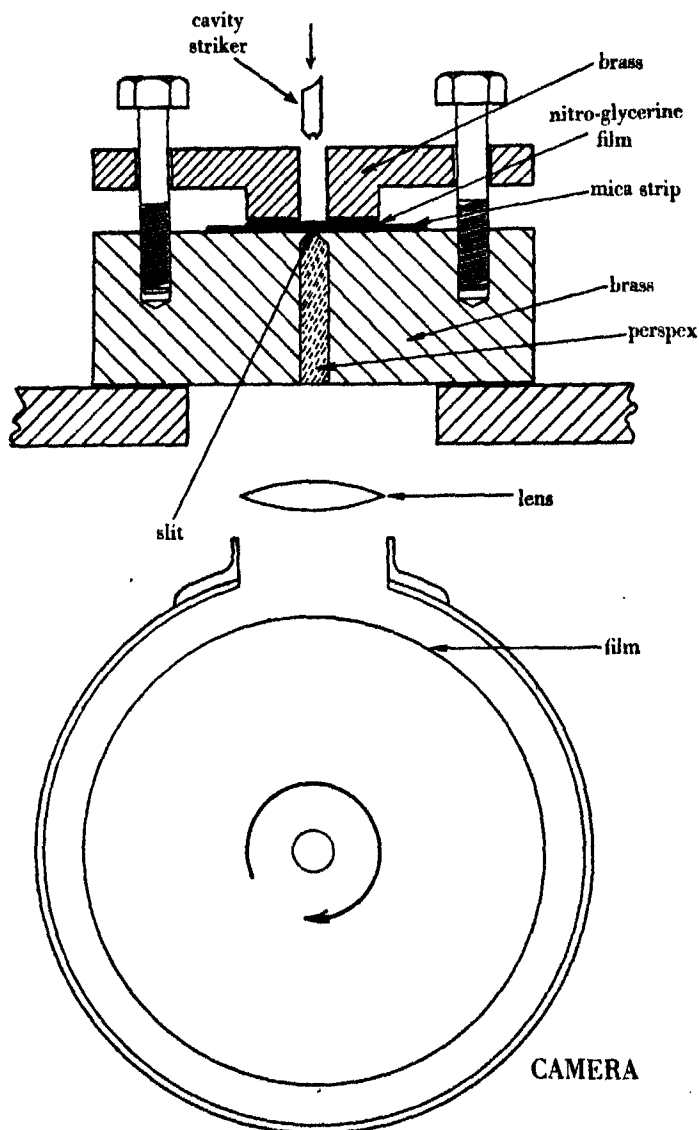


FIGURE 4

By means of this apparatus it was possible to spread out along the length of the moving film some of the successive events shown by the stationary pictures (figure 3, plate 6) and thus determine their chronological order.

Results. Figure 2, plate 5 is a characteristic moving-film picture of cavity impact leading to propagation through a confined film of nitroglycerine, and should be compared with the stationary picture obtained under similar conditions (figure 1*b*). It will be seen that the light of explosion first manifests itself at a point *A* which corresponds to the cavity in the striker, but it is not until the camera drum has moved through a distance of *AB* that the main propagation of the explosion outside the cavity begins. *AB* corresponds to a time of $60\ \mu\text{sec}$. There is, therefore, a time interval of $60\ \mu\text{sec}$. between the first appearance of light in the cavity and the beginning of propagation which, it will be noted, begins at a high velocity $> 1000\ \text{m./sec}$. Other experiments under similar conditions, but with shorter lengths of explosive layer, have confirmed the general nature of the phenomenon. The bright light seen inside the cavity in the stationary picture is, therefore, the first event, and it is due to a prolonged, slow, and essentially gentle burning of the explosive within the cavity. The point *B*, where the propagation of the main explosion begins, evidently corresponds to the eruption from the cavity which causes the channels observed in the rim (see figure 3, plate 6).

It will be seen that the sudden transformation from the first to the second stage of propagation which occurs at some distance from the cavity and which is shown plainly both in stationary photographs and in the blast patterns, corresponds to no obvious discontinuity in the slope of the moving-film pictures. In the propagation photographed in figure 2, plate 5, the transformation to heavy blasting as shown by the pattern on the surfaces was at about the point *C*. In figure 2 there is no very obvious change in the velocity of propagation at this point. It is possible that more detailed measurements with a higher speed camera might reveal some change, and experiments to investigate this are being carried out.*

Rough calculations of the velocity of propagation of the explosion in the thin layer of nitroglycerine have been made from a few pictures similar to figure 2, plate 5. The value obtained is about $1000\ \text{m./sec}$. A somewhat more accurate value has been obtained from figure 6, plate 6, where the initiation of the explosion occurred outside the picture. This gives $1350\ \text{m./sec}$. These films of explosive are, of course, very thin (about $0.1\ \text{mm}$. or less).

The explosion could also be initiated by a spark instead of by cavity impact. The apparatus was similar to that shown in figure 4, except that the top plate had a small hole drilled in it, down which an insulated electrode could pass. The spark was passed by discharging a condenser ($250\ \text{V}$, $4\text{--}7\ \mu\text{F}$). Figure 5, plate 6, is a picture

* *Note added 20 February 1946.* Later experiments with a faster camera by Vines and Mulcahy (1946) show clearly that there is a change in velocity at this point. Over the first stage the velocity is about $400\ \text{m./sec}$., and this suddenly transforms to the second stage of heavy blasting at the point *C*, where the velocity of propagation rises to about $1800\ \text{m./sec}$.

obtained under these conditions. It will be observed that there is again an appreciable time interval (about $30\ \mu\text{sec.}$) between the passage of the spark and the beginning of propagation of the explosion. Other experiments have confirmed the existence of this time interval which may be due to the initial localized burning of nitroglycerine in the spark hole. Under some experimental conditions, the blast patterns show alternate rings of unmarked metal followed by heavily blasted metal. These are very apparent if one of the confining surfaces is a thin sheet instead of a massive block. The pattern may be due to standing waves set up in the metal or to an alternate growth and dying away of the explosion intensity. At the time of writing, however, the origin of these was not certain, and further work is in progress.

Detailed examination of events inside the cavity

Sequence of events. A modification of this technique was used so that attention could be confined to events occurring within the cavity itself. The apparatus (see figure 7) was arranged so that a hemispherical striker with a cavity in it was allowed to fall into an unconfined film or lens of nitroglycerine on a mica sheet or thin glass plate. A photographic plate was placed under the mica sheet for the stationary pictures. For the moving-film pictures, a slit was used which, as before, was set at right angles to the direction of the rotation of the camera drum. The cavity is only 0.4 mm. in radius and the photographs given are magnifications of the light effects.

Some characteristic *stationary* photographs of the light within the cavity are shown in figure 3, plate 6. Figure 3*a* is interesting, since it shows clearly the uneven illumination and incomplete combustion which may occur within the cavity. In this case it is obvious that the explosion has not propagated beyond the cavity. Figure 3*b* is an example of more complete illumination within the cavity itself. Figure 3*c* and *d* show typical examples of eruption from the cavity. The flame blasts through the wall of the cavity into the surrounding explosive. Since the surrounding explosive in this case is not confined, it is blown away and the propagation does not proceed far. If the surrounding explosive were confined as a thin film, the eruption would cause propagation of the explosion throughout the film (cf. figure 1*b*, plate 5). A photograph of the imprint of the cavity on a lead-plated brass anvil and of the blast marks produced is given in figure 3*e*, plate 6. The lead film within the cavity is little damaged, but the eruption under the cavity rim is clearly seen. The resemblance between the blast marks and the light photographs of figures 3*c* and *d* is obvious. Figure 3*f* is a good example of localized burning within the cavity obtained with blasting gelatine. Again the burning is confined to a small region of the cavity and, in this case, the light photograph provided the only evidence that initiation had occurred.

Corresponding enlarged photographs of these events made on a *moving-film* camera are given in figures 8–16, plate 7. As previously mentioned, this apparatus enables the sequence of events, and the velocity of the flame propagation and its

duration to be determined. The direction of the movement of the film is shown on the figure by the arrow, and the time scale is given in microseconds.

The light of the 'flame' within the cavity first appears at *A* and then spreads into the cavity for a distance which is approximately equal to the diameter of the cavity *BB*. The burning proceeds for an additional time equivalent to the film

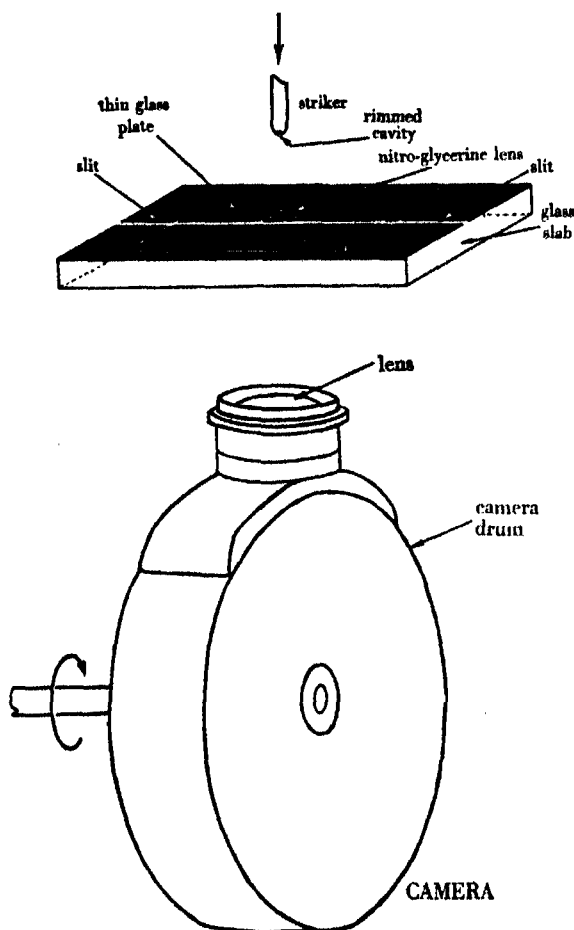


FIGURE 7

movement *BC* without spreading further. At this point *C* the explosion breaks out of the cavity into the surrounding explosive, in which it is propagated for a short distance at high velocity. This is shown by the horizontal direction of the line *CD* which is due to the light from the explosion outside the cavity. The time from the first appearance of the light in the cavity to the eruption (i.e. *AC*) is about $40 \mu\text{sec}$. The burning within the cavity still proceeds until the drum has reached point *E*, which represents in the picture $170 \mu\text{sec}$. after the first appearance of light. A similar

trace is given in figure 9. These moving-film pictures should be compared with the stationary pictures (figures 3c,d, plate 6) which also show the eruption.

A large number of similar experiments showed this sequence of events to be general. At times, however, the burning merely flared up and died down inside the cavity without breaking through (see figures 10, 16). In these cases, only a feeble noise was heard and the behaviour is clearly similar to that shown in the stationary pictures, figures 3a and b, plate 6.

The burning usually continues in the cavity for some time after the eruption (figures 8, 9, 14 and 15), but at times it is extinguished at the moment the eruption occurs (see figures 11, 12). Occasionally, instead of being extinguished at the eruption, the intensity is diminished and it flares up again before finally going out. Figure 13 is an extreme example of this.

Duration of burning. An extensive series of observations has shown that under similar experimental conditions there is some variation in the duration of the burning inside the cavity and in the time at which the eruption occurs. These times, however, are all of the same order of magnitude and are comparatively little affected by variations in the size of the cavity or in the nature of the striker and anvil. In a series of ten experiments with nitroglycerine, the time from the first appearance of light in the cavity to the eruption varied from 41 to 94 $\mu\text{sec.}$ (average 72 $\mu\text{sec.}$). The total time for which the cavity burns depends, of course, upon whether it is extinguished at the eruption or persists afterwards, and variations from 57 to 163 $\mu\text{sec.}$ (average 103 $\mu\text{sec.}$) were observed. These figures refer to a brass striker weighing 60 g. falling 20 cm. on to a glass surface, the diameter of the cavity being 0.7 mm. and its depth also 0.7 mm.

The time from the first appearance of light to the eruption, and the total duration of burning in the cavity, have been measured in a considerable number of experiments in which the conditions of impact were varied. The results show that both these times are substantially independent of whether the cavity strikes the anvil squarely or is slightly inclined. The times, for solutions of nitroglycerine in toluene, are also similar to those for nitroglycerine itself. However, with a soft anvil, the time from the beginning of detectable light to the eruption is somewhat lower than with glass anvils—average 48 $\mu\text{sec.}$; and again a similar though less marked effect is observed when softer (antimonial lead) strikers (Brinell hardness 14) are used—average 55 $\mu\text{sec.}$ With impacts of higher energy than 1200 g.cm., e.g. at 3800 g.cm., it appears that there is a reduction in the total burning time within the cavity—average 60 $\mu\text{sec.}$, although the results cannot be regarded as conclusive. When large cavities are used (diameter 1.4 mm., depth 1.1 mm.) there is a tendency for a longer total burning time within the cavity—average 174 $\mu\text{sec.}$

There is frequently a diminution of light intensity within the cavity immediately after the eruption, but this diminution is transient. The effect is especially marked when antimonial lead strikers are used. In several of such cases the light has been observed to be almost extinguished after the eruption and flare up again some 175 $\mu\text{sec.}$ later (see figure 13, plate 7).

PART II. MEASUREMENTS OF THE PERIOD OF TIME BETWEEN CONTACT OF STRIKER AND ANVIL AND THE APPEARANCE OF LIGHT IN THE CAVITY

The experiments just described show that, after initiation, a burning process takes place for some time inside the cavity before it breaks through into the explosive outside. There still remains the question of the time at which the initiation takes place, i.e. whether before, at or after the first instant of impact between striker and anvil. Previous conductivity experiments (Bowden *et al.* 1943*a*) have shown that initiation occurs at a time which is close to the first instant of solid contact. A more accurate determination of the time of initiation has been made with the aid of the rotating drum camera. The apparatus was similar to that described above, except that in order to ensure rigidity the explosive was put directly on the slit. A comparatively thick layer of aluminium was sputtered on to a glass slab so as to render it opaque, and a narrow rectangle 0.5 mm. wide by 15 cm. long cut out of the aluminium layer so as to form a transparent slit. The surface of the slab was then given a further light aluminizing so that the slit, while remaining optically transparent, was at the same time electrically conducting. A neon discharge lamp was set above the slit beside the explosive. A thyratron circuit was arranged to give a single flash of the lamp at the instant when the metal striker made electrical contact with the aluminized surface. The width of the slit was approximately equal to the diameter of the cavity which fell symmetrically on it.

The device thus caused a flash of light to appear on the film at the precise instant when the striker made contact. In this way, any time lag between contact and the appearance of detectable light within the cavity appears on the film as a displacement of the cavity light from the first point of appearance of light from the lamp. The absence of any instrumental time lag between contact of hammer and anvil and the flash of the neon lamp was checked in separate experiments by allowing a spark passing from the hammer to the slit to trip the circuit. In this case the light from the spark and that from the lamp appeared side by side on the film.

The results with the cavity explosion show that the first point of light in the cavity appears on the film either simultaneously with the beginning of the flash from the neon lamp (dotted line in figures 17–19, plate 8) or at some short time (0–30 $\mu\text{sec.}$) subsequently. In six of the twenty experiments performed, light from the cavity appeared on the film at the instant of contact within the experimental error ($\pm 5 \mu\text{sec.}$). In the remaining fourteen experiments, the first detectable light appeared, on the average, 18 $\mu\text{sec.}$ after the first moment of solid contact.

Mechanism of action of the cavity after initiation

The evidence presented has led to the view that after initiation within the cavity, by adiabatic compression of the entrapped gas bubble, the following sequence of events takes place.

(a) At initiation a pin-point of light appears (frequently near the rim of the cavity) and a slow burning of the explosive begins and spreads within the cavity. If it is assumed that the pointed head of the trace of light from the cavity (e.g. figures 8, 10, plate 7) is due to the compounding of the velocity of the flame in the direction of the slit and the velocity of the drum at right angles to it, then it is possible to calculate very approximately the velocity of propagation of the flame front. Values ranging from 6 to 25 m./sec. have been obtained. Taking into account that the flame front will not in general be travelling parallel to the slit, we may say that, if the initial assumption be correct, the flame within the cavity is propagated at a velocity of about 20 m./sec.

(b) After the flame front has moved in the cavity for some distance, leaving behind it a region of material still burning, there comes a point when the pressure developed by the products of combustion is great enough to burst out of the sealed reaction chamber formed by the cavity. This is the point of eruption and represents the first point at which any flame or explosion process may be detected outside the cavity. It may occur when only part of the nitroglycerine in the cavity has been consumed. When the reaction chamber has softer walls, i.e. with anvil or hammer of softer material, the pressure necessary to cause the eruption will be less and the time from the initiation to eruption should be shorter. This also has been observed.

(c) After the eruption from the cavity, an explosion may be propagated a short distance through the unconfined nitroglycerine about the cavity with high velocity (about 1000 m./sec. or greater). If the explosive outside the cavity is confined as a film between two surfaces, the propagation spreads through the whole of the film.

(d) After the eruption, the burning process within the cavity sometimes undergoes no immediate alteration, but it may suffer a diminution in intensity or even go out. These latter effects are probably due to the sudden release of pressure. In some cases the diminished or even extinct burning process may restore itself. This may be associated with resealing of the cavity as the material of the striker yields further under the impact. The burning within the cavity with impact of 1200 g.cm. normally continues for some 70–120 μ sec. after the eruption occurs.

The electrical measurement of the period of contact in cavity explosions

The photographic observations described above have been amplified by an electrical measurement of the conductivity between the impacting surfaces, from which the period of impact may be obtained, and by measurements with a piezo-electric crystal of the time of beginning of propagation of the shock wave of the explosion.

The total time of contact. The period of contact between striker and anvil was determined by means of a cathode-ray oscillograph. With this method it is possible to study the rapidly changing electrical resistances during impact. Before the two surfaces touch, the electrical resistance between them is infinite. As they touch, this resistance decreases suddenly to a very low value and then increases to an

infinite value again as the surfaces separate. The general shape of the curves obtained is illustrated in figure 20 and it is evident that plastic deformation does occur. When a brass cavity striker weighing 35 g. is allowed to fall 10 cm. (energy 350 g.cm.) on to a clean brass anvil, the average period of contact was found to be 160 μ sec. (130–180 μ sec.). A film of nitroglycerine which does not explode has no measurable effect on the period of contact or on the minimum resistance. When explosion occurred, the period of contact was somewhat lower—average 130 μ sec.—presumably because of the lifting effect on the hammer, due to the explosion.

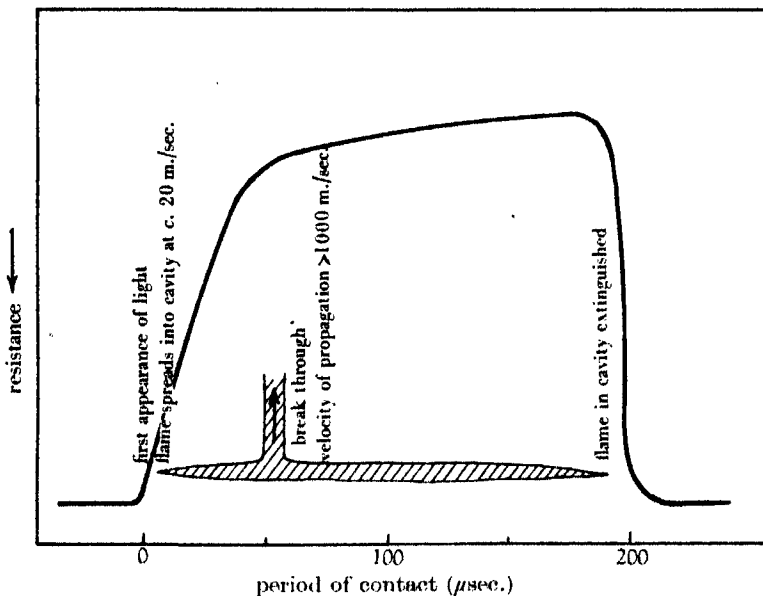


FIGURE 20. Shows diagrammatically the period of contact of brass cavity strikers on a brass anvil.

Similar experiments were also carried out using thin glass anvils covered with a fine conducting layer of aluminium. The period of contact was approximately the same as for brass surfaces (140–200 μ sec.).

When flat brass strikers 1.5 cm. diameter were used instead of the hemispherical cavity strikers, the period of contact between the striker and anvil was somewhat increased. With a striker weighing 180 g. falling 30 cm. the period of contact was about 300 μ sec.

A recording of the shock of the explosion. A 400 kc. 'scale-of-two' counter chronometer has been developed by J. S. Courtney-Pratt and A. E. Ferguson which can measure time intervals between any two electrical impulses to the nearest 2.5 μ sec. This instrument has been used to measure the time at which the shock from the explosion commences relative to the first instant of metallic contact.

The apparatus used consisted of a flat brass anvil resting on a small quartz crystal embedded in a brass block. The striker was allowed to fall on to a lens of

nitroglycerine placed on the brass anvil directly above the crystal. When contact occurred between the striker and anvil, the 400 kc. counter was set in operation and this was switched off as soon as the shock wave from the explosion reached the crystal. By this means it was possible to measure the time interval between the metallic contact and the arrival of shock wave at the crystal detector. Due allowance for the time taken for the shock wave to travel through the brass anvil was made. The results showed some variation, but it was clear that in the case of cavity strikers there was no prolonged time lag from contact to initiation, and the values are of the same order of magnitude as the time lags obtained on the rotating drum camera.

When a *flat* striker which did not contain a cavity (180 g. falling 20 cm.) was allowed to fall on to the nitroglycerine distributed as a number of droplets, the shock wave in general operated the chronometer before actual metallic contact between the surfaces took place. It appeared therefore that under these conditions explosion could occur before the metal surfaces came into contact.

PART III. MEASUREMENT OF VELOCITY OF FLOW OF LIQUID STRIPS AND THE TIME LAG BETWEEN COALESCENCE AND EXPLOSION

The explosion of nitroglycerine when distributed as two parallel strips has been discussed in the previous paper. It was of interest to ascertain at what point of time the explosion is initiated relative to the impact of the flat hammer on the strips. Experiments to this end were carried out with a rotating drum camera. The apparatus is shown in figure 21. Two strips *F* of nitroglycerine were placed 0.75 mm. apart on a thin glass plate *P* which covered a slit cut out in a metal block. The nitroglycerine was rendered opaque by dissolving about 0.1 % of dyestuff (methyl violet) in it. The slit was arranged at the focus of the lens *K* parallel to the axis of rotation of the drum *D* which carried the film. In one series of experiments the strips were arranged across the slit (figure 22*a*), in a second series parallel to the slit and directly over it (figure 22*b*). A beam of light from lamp *L* and condenser *C* was passed vertically through the slit. With the arrangement shown in figure 22*a*, two shadows corresponding to the coloured strips were cast on the rotating film. The strips were struck with a cylindrical glass hammer *H*, 2.5 cm. in diameter, the bottom surface of which had been arranged to be accurately flat with respect to the glass plate *P*. The hammer weighed 240 g. and fell through about 40 cm. Shortly before impact, an electromagnetic device opened the camera shutter which remained open for approximately one revolution of the drum. The amount of explosive used was in each case about 0.05–0.1 c.c.

With the arrangement shown in figure 22*a*, the velocity of the spreading of the two strips under the impact can be measured by the rate at which the two shadows expand outwards and towards each other on the film. And further, the instant at which explosion begins is shown on the film as the point of first appearance of the light from the explosion relative to the beginning (or end) of the movement of the



(b)



(a)

FIGURE 1



FIGURE 2

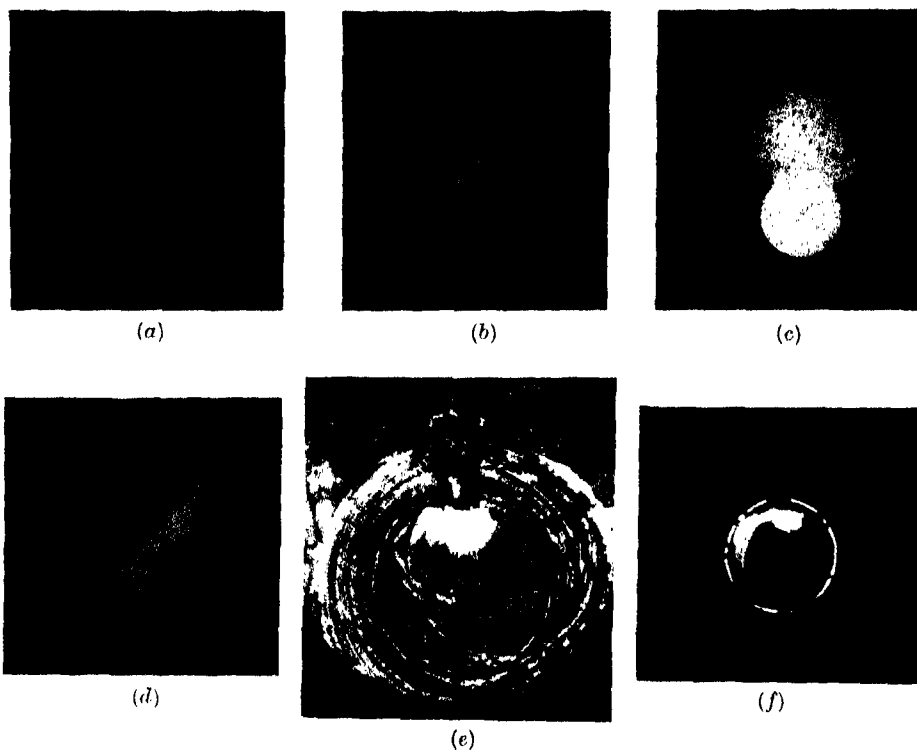


FIGURE 3



FIGURE 5



FIGURE 6

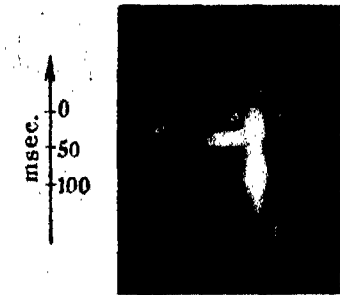


FIGURE 8

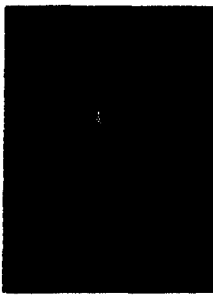


FIGURE 9

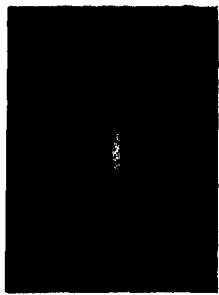


FIGURE 10

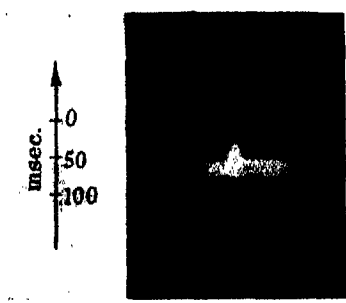


FIGURE 11



FIGURE 12

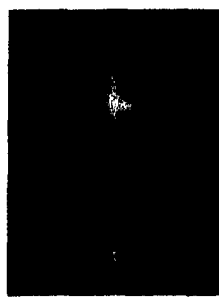


FIGURE 13

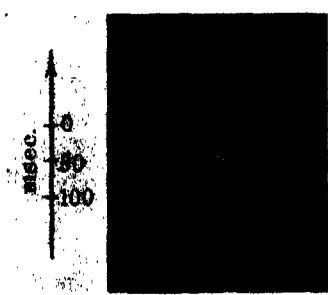


FIGURE 14

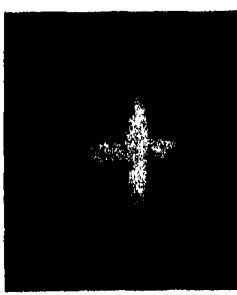


FIGURE 15



FIGURE 16

strips. Figure 23 *a*, *b* and *c*, plate 9 show typical pictures obtained: *a* in which no explosion took place, and *b* and *c* in which initiation was effected. It will be seen in figure 23*a* that motion of the liquid under impact begins at *A* and *A'*; the strips are then squashed towards each other and outwards. The velocity of the outward spreading is given by the slope of the lines *AB* and *A'B'* (in this case 86 m./sec.). At *C* the light between the strips is extinguished; this, therefore, marks the point of coalescence. In some cases an approximate value for the velocity of coalescence could be ascertained from the slope of the lines through *C*.

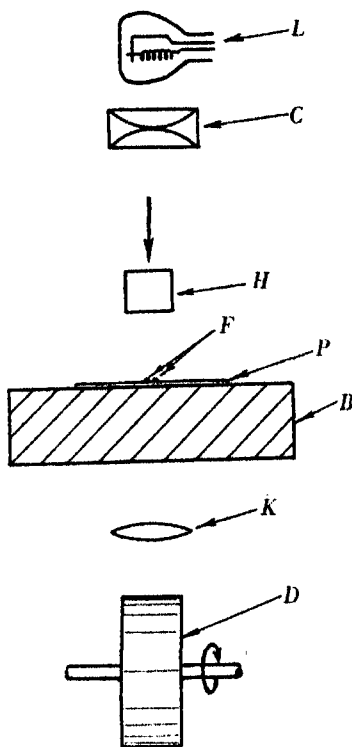


FIGURE 21. Arrangement of optical apparatus for strip experiments with drum camera; showing lamp *L*, condensing lens *C*, transparent glass striker *H*, strips of nitroglycerine *F*, glass plate *P*, steel anvil with slit (not shown) *B*, camera lens *K* and rotating drum *D*.

In figure 23*b*, explosion has resulted. It will be seen that again the squashing of the strips begins at *A* and at *A'*, and coalescence occurs at *C*. The explosion does not begin until the camera film has moved on from *C* to the point *D*, where the first explosion light is recorded. This distance corresponds to 32 μ sec.

In figure 23*c* there is no apparent time lag between coalescence and the initiation.

The time lags observed between the coalescence of the strips and the explosion with the arrangement as in figure 22*a* for a series of experiments with a 235 g. striker falling 40 cm. are: 0, 12, 16, 22, 24, 32 μ sec. With the same conditions of

impact the values obtained for the average velocity of outward flow of the parallel strips of nitroglycerine spreading under the impact are: 65, 72*, 74, 74, 75, 84*, 86, 105, 130* m./sec. Initiation occurred only in those experiments marked with an asterisk. It will be seen that, in these cases, the velocity of spreading is not appreciably different from the experiments in which no explosion took place.

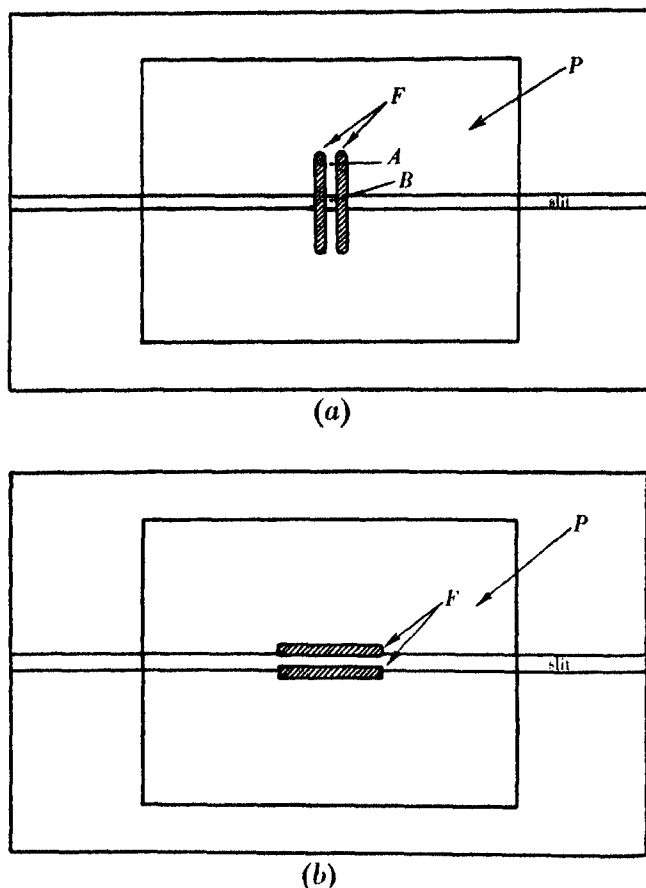


FIGURE 22. (a), Strips F on glass plate P placed so that they lie at right angles to slit.
(b), Strips F on glass plate P parallel to slit.

The results for the time between coalescence and explosion indicate that in some cases a time lag is observed from the instant of coalescence of the strips to the beginning of light from the explosion. It would seem at first sight that this could possibly be due to initiation of the explosion at some point not photographed, e.g. at the point A in figure 22*a*. In this case the explosion would have to propagate from A to B before it would be recorded on the film. This would normally take only about $5 \mu\text{sec.}$, assuming a velocity of propagation of 400 m./sec.

If it be assumed that initiation occurs at or near the junction of the two strips, the arrangement shown in figure 22*b*, wherein the whole length of the junction is affected over the slit, precludes the possibility of an initiation invisible to the camera. In this case the coalescence is marked by the extinction of the light passing through the slit. A number of such experiments was tried. Figure 24*a*, plate 9, shows such an experiment in which no explosion took place. The film moved in the direction of the arrow; *AA'* marks the coalescence of the strips. (The uneven edge of the shadow is due to the fact that the strips are not perfectly straight, so that the coalescence has taken place at some points slightly earlier than at others.) Figure 24*b*, plate 9, is a similar experiment in which explosion was initiated. Coalescence is shown by the beginning of the shadow *AA'*, but the film has moved on a short distance before explosion begins at point *B*. Thus, in this case, there is a real time lag of 16 μ sec. between coalescence and the beginning of the explosion (or, more strictly, the beginning of light emission). In the experiment shown in figure 24*b* the strips had coalesced from the beginning over a short length at the left-hand side, and this is shown by the continuous shadow *C*. Figure 24*c*, plate 9, is another such experiment. Here the time lag between the instant of coalescence *A* and the beginning of explosion *B* is less than 8 μ sec. Light is visible between *A* and *B*, which may be attributed to the presence of an entrapped bubble. It is clear that in this case the initiation has begun near the last point of coalescence of the strips; this has been found in a number of cases.

Values of the time lag between coalescence and the beginning of explosion with the arrangement as in figure 22*b* are 8, 10, 10, 10, 16, 28, 56, average 20 μ sec. Again the striker weighed 235 g. and the height of the fall was 40 cm. and these figures are about the same as those obtained with the other arrangement (figure 22*a*). It will be seen that the time lag between coalescence and initiation is short. The average value for the seven experiments is 20 μ sec. Experiments were performed in which only one of the strips contained dissolved dyestuff. The results obtained were similar to those given above.

Initial stages of propagation in thin films after initiation by coalescence of strips

The propagation of the explosion through a thin film of nitroglycerine after it has been initiated by cavity impact has been described earlier. After a period of gentle and prolonged burning within the cavity, the explosion is propagated throughout the surrounding film with a velocity of about 1000 m./sec. The blast patterns on the metal surface suggest that the propagation occurs in two stages. The first stage which may extend radially for a distance of 5 mm. from the point of initiation causes little damage to the surface; after this the explosion causes a marked blasting and depression of the metal.

A study of the rate of propagation of the explosion after it has been initiated by the entrapping of a bubble compressed between coalescing strips has also been made. Some typical results are shown in figure 24*b* and *c*, plate 9. The initial prolonged and gentle burning which takes place inside the cavity before propagation begins

is not observed here. There is, however, evidence for two stages of propagation. There is an initial 'slow' velocity of propagation of about 400 m./sec. (average of nine experiments) which is maintained from the point of initiation for a radius of some 5 mm. and which then passes over to a more rapid velocity of about 1000 m./sec. or greater. This region of 'slow' velocity about the point of initiation corresponds roughly with the unmarked region on the metal anvil described above.

In a recent paper Chariton & Ratner (1943) have reported values for the detonation velocity of nitroglycerine initiated by a detonator in glass tubes of various diameter. They find that in tubes of greater diameter than about 5 mm. the detonation may proceed at either of two velocities, viz. about 1800 m./sec. or about 7500 m./sec. On the other hand, in tubes of less than about 2 mm., the detonation is invariably propagated at the lower velocity (1800 m./sec.). They also record the existence of an unstable propagation at about 700 m./sec. in 3-4 mm. tubes of nitroglycerine. Ratner (1944) has also demonstrated the existence of two definite velocities with nitroglycol and methyl nitrate.

It is interesting to note that the initial velocity of propagation of the explosion is about 400 m./sec., and the velocity of outward spreading of the explosive is about 100 m./sec. The pictures (e.g. figure 24*b*) show that, although spreading of the liquid begins before initiation occurs, the explosion soon overtakes the moving liquid.

Pictures obtained from experiments in which one of the strips did not contain dyestuff show that the liquid is stationary in the region where initiation occurs.

DISCUSSION AND CONCLUSION

These experiments support the conclusion that the high sensitivity of nitroglycerine and other liquid explosives to gentle impact is due to the entrapping of tiny gas bubbles which are heated by adiabatic compression and fire the explosive.

The conductivity measurements on impacting surfaces show that the period of contact between the cavity striker and the anvil is of the order of 200 μ sec. The experiments with the moving-film camera show that initiation first becomes apparent as a tiny pin-point of flame within the cavity and presumably in the immediate vicinity of the gas bubble. This initiation occurs at a time which corresponds to the first moment of contact between the cavity and the anvil, i.e. to the point *A* in figure 8, plate 7, or very shortly (about 20 μ sec.) after it. This tiny flame spreads slowly through the cavity with a velocity which is of the order of 20 m./sec. After a period of time of about 50 μ sec. the flame and hot gases burst through the rim of the cavity and fire the main film. After the eruption the explosive inside the cavity may continue burning, but is finally extinguished when the hammer leaves the anvil. The sequence of events is illustrated diagrammatically in figure 20. The explosion is propagated through the thin film of nitroglycerine with a velocity of about 1000 m./sec.

In the second series of experiments the explosion is initiated by the compression of a tiny gas bubble entrapped between coalescing strips (or droplets) of the

explosive. The photographs show that, under the conditions used in these experiments, the nitroglycerine in the parallel strips flows rapidly (velocity about 100 m./sec.) under the flat impacting surfaces. A small bubble of air is usually entrapped near the last point of coalescence and initiation frequently occurs at this point, although examination after the impact shows that numerous other bubbles are entrapped. Again the initiation begins as a small flame which becomes visible at the moment of coalescence or a short time (about 20 μ sec.) after it. It will be observed that the duration of the average time lag which elapses between coalescence and initiation in these experiments corresponds with that found in the cavity experiments, viz. between the time of contact of hammer with the anvil and the initiation. The coalescence of the strips and the initiation of the explosion can occur when the surfaces are still separated by a macroscopic film of the explosive. The flame first propagates radially outward through the film with a velocity of about 400 m./sec. and after a distance of about 0.5 cm. it transforms into a higher velocity and is propagated at about 1000 m./sec. This transformation would appear to correspond with the change in the blast pattern on the confining surfaces.

These results are all in harmony with the view that the initiation is essentially a thermal one brought about by an adiabatically heated gas bubble. The short time lag between the application of the pressure and the initiation of the explosion shows that the temperature of the tiny bubble must be very high. It was suggested in the last paper that initiation of the nitroglycerine may occur in the vapour phase inside the bubble, and the onset of the explosion and the shortening of the time lag may well be assisted by the high pressure which is known to exist in the liquid during the period of impact.

REFERENCES

- Bowden, F. P., Eirich, F., Ferguson, A. E. & Yoffe, A. 1943*a* *Bull. Coun. Sci. Industr. Res., Aust.*, no. 167.
Bowden, F. P., Eirich, F., Mulcahy, M. F. R., Vines, R. G. & Yoffe, A. 1943*b* *Bull. Coun. Sci. Industr. Res., Aust.*, no. 173.
Chariton, J. B. & Ratner, S. B. 1943 *C.R. Acad. Sci. U.R.S.S.* **41**, 293.
Ratner, S. B. 1944 *C.R. Acad. Sci. U.R.S.S.* **42**, 265.
Vines, R. G. & Mulcahy, M. F. R. 1946 *Nature*, **157**, 625.

DESCRIPTION OF PLATES

Plate 5

FIGURE 1. *a*. Pattern on brass produced by thin film of nitroglycerine confined between two flat surfaces. Explosion initiated by rimmed cavity impact. Observe central unmarked area about imprint of cavity, and transformation to heavy blasting at *C*. ($\times 2.8$.) *b*. Light pattern under conditions similar to those of figure 1*a*. Observe (i) bright illumination within cavity, (ii) region of weak illumination around cavity and abrupt transformation to second stage of brighter illumination. ($\times 2.8$.)

FIGURE 2. Light trace (moving film) obtained during propagation of explosion initiated by cavity impact at *A*. Conditions similar to those of figure 1*a* and *b* (cf. stationary picture, figure 1*b*). Cavity burns for time *AB* before propagation occurs.

Plate 6

Explosion initiated by cavity impact. Luminous and blast effects

FIGURE 3. *a.* Rimless cavity impact. Photographic plate only partially developed. Note non-uniform illumination within the cavity. ($\times 18$.) *b.* Rimless cavity impact. Complete development of plate. Note uniform illumination of the cavity and absence of any exterior luminous effects. ($\times 13$.) *c.* Rimless cavity impact. Strong eruption from cavity. ($\times 11$.) *d.* Nitroglycol light pattern. Observe pronounced eruption from cavity. ($\times 10$.) *e.* Brass anvil (lead plated) after high energy cavity impact. Note pronounced eruption channel breaking through wall of cavity. ($\times 23$.) *f.* Blasting gelatine. Light pattern showing uneven distribution of luminous effects within cavity—outline of cavity indicated by dotted line (cf. figure 3*a*). ($\times 16$.)

FIGURE 5. Spark initiation. Light trace obtained during propagation of explosion over 10 cm. Initiation by spark at *A*. Observe time lag between passage of spark and beginning of propagation. Diameter of deflagration area corresponds to distance *AP* (cf. figure 2).

FIGURE 6. Moving-film photograph of propagation through thin film of nitroglycerine. Point of initiation not shown.

Plate 7

Light traces from cavity impacts with identical time scale ($\mu\text{sec.}$)
but varying lateral magnifications

FIGURE 8. Light traces from rimmed cavity impact. First appearance of light (in cavity) at *A*. Break-out at *C* and propagation of explosion into unconfined nitroglycerine along *CD*. Continuation of burning within cavity to *E*.

FIGURE 9. Light trace obtained under conditions similar to those of figure 8.

FIGURE 10. Light trace obtained under conditions similar to those of figure 8. No break-out in this case.

FIGURE 11. Light trace from higher energy cavity impact. No light visible in cavity after break-out.

FIGURE 12. Light trace obtained under conditions similar to those of figure 8. In this case the detonation was violent. No light visible in cavity after break-out.

FIGURE 13. Light trace using antimonial lead striker. Observe reappearance of light some time after break-out.

FIGURES 14 and 15. Light traces under conditions similar to those of figure 8, showing well-defined break-out and continuation of burning within the cavity.

FIGURE 16. Light trace obtained with nitroglycol. Break-out absent (cf. figure 10).

Plate 8

FIGURE 17. Light trace obtained from rimmed cavity explosion (right) and neon flash (left). White dotted line indicates the instant of metallic contact.

FIGURE 18. As for figure 17.

FIGURE 19. Two independent experiments similar to figure 17.

Plate 9

(Pictures obtained with apparatus shown in figure 21)

FIGURE 23. *a.* Spreading of strips arranged as in figure 22*a*; no explosion occurred in this experiment. Motion begins at *A* and *A'*; strips spread outward with velocity given by slopes *AB* and *A'B'*. *C* is the point of coalescence. *b.* Strips arranged as for 23*a*. Explosion occurs at *D* some 30 μ sec. after coalescence. *c.* Similar to 23*b*. In this case there is no apparent time lag between coalescence and initiation.

FIGURE 24. *a.* Coalescence of strips arranged as in figure 22*b*. No explosion occurred. Edge of shadow *AA'* represents the instant of coalescence of strips. *b.* Strips arranged as for 24*a*. Initiation of explosion occurs at *B* 16 μ sec. after coalescence. (Strips had coalesced at left-hand end *C*, prior to experiment.) *c.* Similar to 24*b*. Initiation of explosion at *B*, very shortly after coalescence at *A*. Initiation occurs near last point of coalescence of strips.

Hot spots on rubbing surfaces and the detonation of explosives by friction

By F. P. BOWDEN, M. A. STONE AND G. K. TUDOR, *Tribophysics Section,
Council for Scientific and Industrial Research, University of Melbourne*

(Communicated by Sir David Rivett, F.R.S., and Sir Robert Robertson, F.R.S.—
Received 12 January 1946—Read 9 May 1946)

[Plates 10 and 11]

The local temperature at the surface of sliding solids may reach a high value even under comparatively gentle conditions of sliding. With metals the surface temperature can be measured by using two different metals as a thermocouple; the temperature depends on the load, speed and thermal conductivity of the metals. With non-metals, which are poor thermal conductors, the high temperatures occur even more readily. The thermoelectric method cannot be used with non-conductors, but if one or both of the surfaces is of transparent material such as glass or quartz, the hot spots may be observed visually or recorded photographically. Experiments with surfaces of different melting-points show that the hot spots first become visible when the temperature is about 520–570° C. Hot spots of this temperature may occur on glass or quartz when the load is about 1000 g. and the sliding speed as low as 1 or 2 ft./sec. These local surface temperatures play an important part in the chemical decomposition which accompanies the rubbing of solids. Experiments show that the detonation of a liquid explosive, such as nitroglycerine, by friction is due to the development of local high temperatures on the surface of the sliding solids. Under the conditions of these experiments the temperature of the hot spots necessary to cause the explosion of nitroglycerine is about 480° C.

INTRODUCTION

It is well known that many explosives can be detonated by friction and a number of practical tests has been devised for measuring the friction sensitivity. One of these consists of a swinging pendulum which strikes the explosive a glancing blow. Another is carried out by allowing a weight with a curved end to slide down an

incline and strike the explosive a glancing blow. Although tests of this type have been used extensively to assess the relative sensitivity of different explosives from the point of view of safety in manipulation and manufacture, little work has been published on the mechanism of initiation by friction. This is particularly true for liquid explosives. Although it is well known that nitroglycerine, for example, may be detonated by friction, the mechanism of this and the conditions under which it can occur are by no means understood. Earlier work (Bowden & Ridler 1936) has shown that, when surfaces are rubbed together, the local temperature at the points of sliding contact may reach a high value. The important part which these local high temperatures play, both in physical processes such as polishing and in the chemical decomposition which accompanies the rubbing and friction of solids, has been discussed elsewhere (Bowden & Hughes 1937). We might expect that these local frictional hot spots could play an important part in the initiation of explosive reactions (see Bernal 1938). It has also been suggested (see, for example, Taylor & Weale 1938) that initiation is a tribochemical effect and is brought about by a direct mechanical activation of the explosive molecules. In the case of a liquid or of an explosive which is liquefied by impact or by rubbing, initiation might be attributed to the very rapid flow and shearing of the explosive or by a viscous heating of the liquid to a temperature above its ignition point. It is clear that these theories are not mutually exclusive and that a combination of factors such as high temperature and high pressure, or high rate of flow and high pressure may be responsible. It is equally true, of course, that the mechanisms of initiation of different explosives may be different and that the excitation of a sensitive primary explosive such as mercury fulminate may be brought about in a very different way from that of a comparatively insensitive high explosive such as nitroglycerine.

Some previous experiments (Bowden 1940) with heavy impact on nitroglycerine between metal surfaces in the presence of hard grit particles have indicated that local hot spots on the grit particles may initiate the explosion. The following paper describes experiments which were carried out to determine how far the frictional high temperatures produced on rubbing surfaces of metals and other solids were responsible for the explosion of nitroglycerine.

When metal surfaces were used, the surface temperature was measured by using the surfaces themselves as a thermometer; that is, by placing two different metals in contact so that they form a thermocouple, and measuring the electromotive force generated on sliding. The general arrangement was similar to that used earlier except that improved electrical circuits were used and the electromotive force was measured on a cathode-ray tube. This enables very rapid fluctuations in the surface temperature of the sliding metals to be recorded.

Since the temperature depends on the thermal conductivity we should expect that high temperatures would be reached much more readily on non-conducting solids such as glass and quartz or on a solid explosive. Obviously the thermoelectric method cannot be used with these solids, but it is possible to show the existence of these hot spots by visual means. If one or both of the surfaces is made

of transparent material, such as glass or quartz, and the apparatus so arranged that a clear image of the rubbing surfaces can be seen, it is found that, when sliding starts, a number of tiny stars of light appear at the interface between the rubbing surfaces. It is clear that they correspond to small hot spots on the surface. The method is not quantitative but—by making one of the surfaces of metal, or metallic alloy of known melting-point, and sliding it on a transparent non-conductor—it is possible to fix approximately the temperature at which hot spots first become visible.

Both these methods have been used to study the temperature of sliding surfaces which are covered with films of nitroglycerine or other liquid explosives. The results provide direct evidence for the view that initiation is caused by frictional hot spots on the sliding surfaces. Under the condition of these experiments, the temperature of the hot spots necessary to cause explosion of nitroglycerine is about 480° C. Temperatures of this order readily occur on the surface of sliding solids under very moderate conditions of load and speed.

After a description of the apparatus used, the first part of the paper describes the visual observation of hot spots, the second the initiation of the explosion of nitroglycerine between the rubbing surfaces and the third a simultaneous measurement of the surface temperature and of the incidence of explosion.

EXPERIMENTAL

Friction apparatus

The friction apparatus was similar to that used in earlier work and consisted essentially of a lower surface in the form of a horizontal circular disk which could be rotated at various speeds, and an upper surface in the form of a small slider which rested under a known load on the rotating surface. The lower disk, which was 25 cm. in diameter and about 2 cm. thick, was made of steel or of glass and was driven by a series of pulleys so that peripheral speeds varying from 50 to 450 cm./sec. could be obtained. The upper surface was usually in the form of a cylindrical slider 0.7 cm. in diameter running with its flat end in contact with the surface of the rotating disk. The load between the surfaces could be varied between 1 and 10 kg.

The lower surface was prepared on a wet lead lap using grade FF carborundum and washed under running water. It was dried by spinning for a few seconds in the air and the measurements were made immediately (care being taken to keep the surfaces grease-free). If nitroglycerine or other liquid films were used, they were added as soon as the surface was dry. The liquid films were prevented from being thrown off the rotating disk by a small rim of paper on the outside edge.

Electrical apparatus

For most of the electrical measurements of surface temperature a constantan slider rubbing on a steel disk was used. The lower steel disk was maintained at earth potential by a lead through a mercury cup mounted in the centre of the plate.

The thermoelectric voltage generated on sliding was amplified by a direct coupled amplifier with a high impedance input ($10,000\ \Omega$). Two amplifying stages, both using EF 50 pentodes, gave a gain of 72 db. ($\times 4000$) with a flat response up to about 150 kcy./sec. The output of the amplifier was fed directly to the horizontal deflecting plates of a high voltage cathode-ray oscillograph. The cathode-ray tube, type 5LP5, had accelerating potentials of approximately 2000 and 4000 V with respect to the cathode on the second anode and intensifier. The thermoelectric voltage was calibrated by applying a series of steady potentials to the amplifier.

A multiplier photo-tube (931) with a total potential of approximately 1000 V was placed about 12 in. from the slider and within view of the explosion. When an explosion occurred, the light flash of the explosion produced a pulse in the photo-tube and this, amplified by a single EF 50 pentode stage, was used to trigger a strobotron tube 631-P1. This tube was completely enclosed, with the exception of a narrow slit arranged to be at the same level as the oscillograph spot. The flash of the strobotron enabled the instant of explosion to be recorded on the camera simultaneously with the temperature trace. The time lag between the flash of explosion and the strobotron flash was very small, probably of the order of $1\ \mu\text{sec}$. In order to extinguish the strobotron, the flash discharged a $2\ \mu\text{F}$ condenser. The time constant of the condenser and the charging resistance limited the time between flashes to a minimum of approximately $1/150\ \text{sec}$. The circuit used to flash the strobotron is reproduced in figure 1. A small neon tube connected through a $0.15\ \text{M}\Omega$ resistance to the 50-cycle mains and mounted beside the strobotron, provided a convenient time-marking device for the film. This electrical apparatus was designed and constructed by A. E. Ferguson and R. W. Muncey.

PART I. VISUAL OBSERVATION OF HOT SPOTS

In the experiments described here, the lower surface was usually a flat glass disk, and a mirror was mounted beneath it in such a way that the region of contact between the slider and the disk could be clearly seen. In some of the experiments, when a transparent slider was used, the top of the slider itself was optically polished and observations made through it. The lower surface was set in motion at a fixed speed and the load on the slider gradually increased. When the load reached a sufficiently high value, a number of small, dull red luminous points was observed on the surface of the slider. The position of these points changed from instant to instant as the points in intimate contact wore away and new points came into contact. If the load (or speed) was increased the spots became brighter and whiter, corresponding to the higher temperatures reached. (The possibility that these luminous spots are due to some triboluminous effect other than a thermal heating must be considered, but the evidence is against this.) The results in the following sections refer to the conditions under which the first dull red hot spots were observed visually in a completely darkened room. The experiments were carried out for clean surfaces and for surfaces wetted with a mixture of glycerine and water.

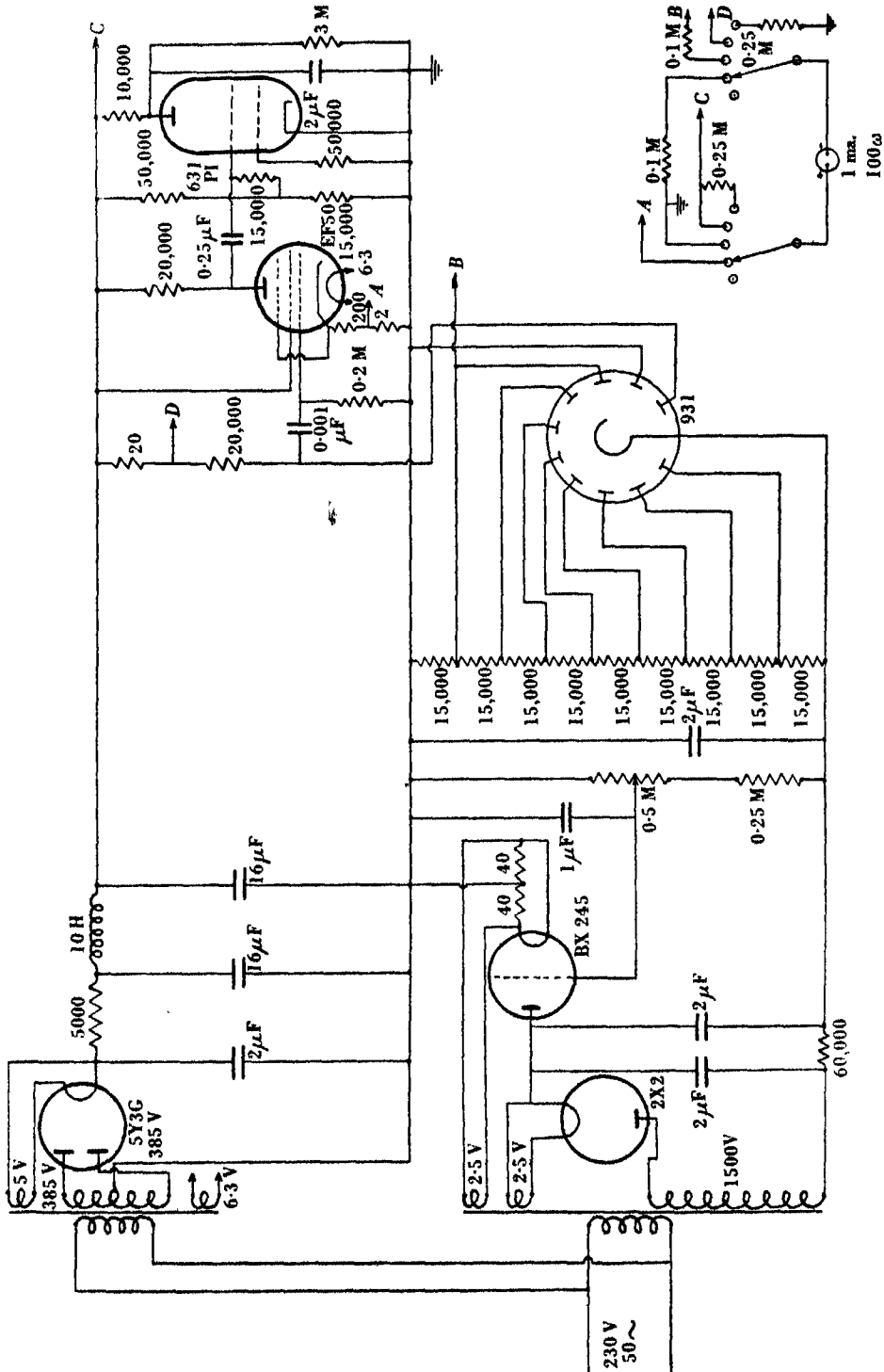


FIGURE 1

Effect of thermal conductivity on the incidence of hot spots

Four hard metals of widely differing thermal conductivity were selected as sliders. The metals used were constantan ($k = 0.05$), steel ($k = 0.10$), nickel ($k = 0.16$) and tungsten ($k = 0.35$ cal. cm.⁻¹ sec.⁻¹ °C⁻¹).

(a) *Clean surfaces.* The results for clean surfaces are shown in figure 2. In this figure, the frictional force at which hot spots occur (for a number of fixed sliding speeds) is plotted against thermal conductivity. It is seen that, in all cases, hot spots occur more readily the lower the thermal conductivity of the slider. This is particularly marked at the lower peripheral speeds. For example, at a surface speed

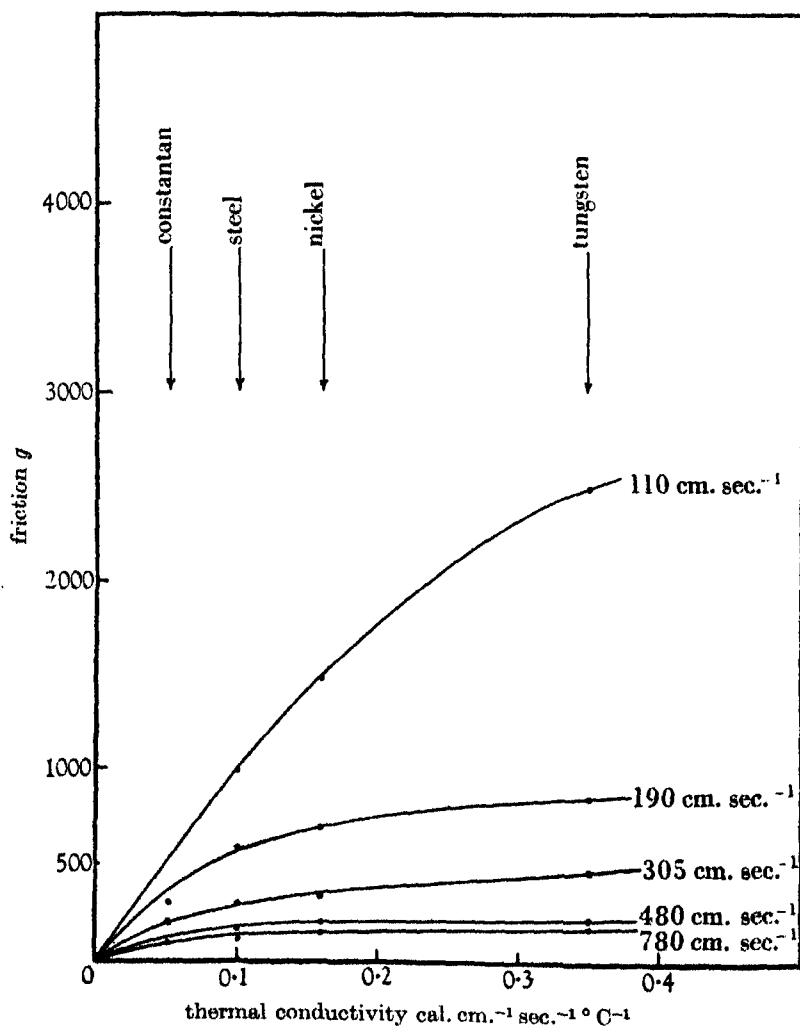


FIGURE 2. The generation of visual hot spots on a clean glass surface. The vertical ordinate gives the frictional force at which hot spots appear for various peripheral speeds, using sliders of constantan, steel, nickel and tungsten.

of 110 cm./sec. a tungsten slider gives hot spots when the frictional force is 2600 g., whilst with a constantan slider hot spots occur when the frictional force is only 350 g.

(b) *Wet surfaces.* The results for surfaces flooded with a mixture of glycerine and water of the same viscosity as nitroglycerine are shown in figure 3. It is seen that the curves are of the same general form as those obtained with clean surfaces. The main difference is that higher frictional forces (six- to sevenfold) are required to produce visible hot spots when the surfaces are flooded with liquid. Although this difference is relatively large, it is evident that the presence of the liquid film is not able to prevent the occurrence of extremely high local temperatures as a result of frictional heating.

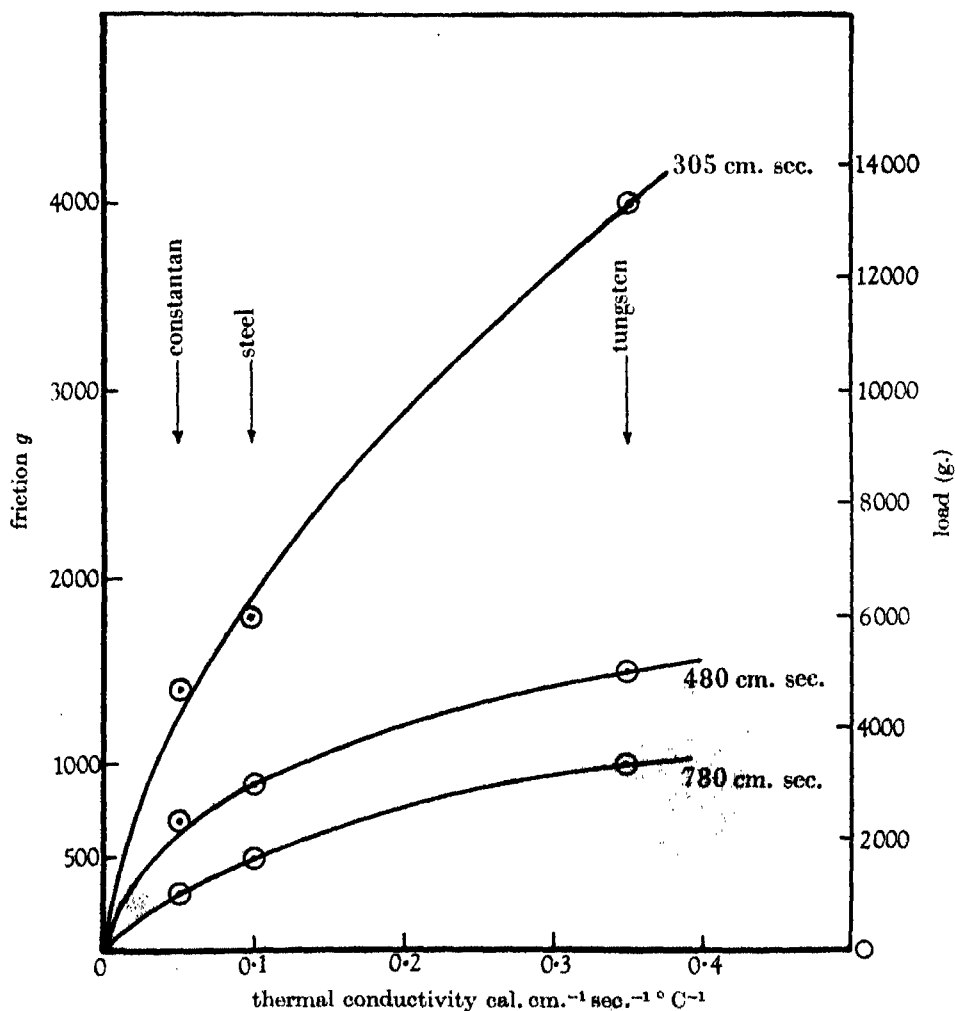


FIGURE 3. The generation of visual hot spots on a glass surface covered with a mixture of glycerine and water. The results are similar to those in figure 2, except that the frictional force necessary to produce visual hot spots is six to seven times higher.

Temperature of incidence of visible hot spots

By using sliders of different melting-point, it is possible to fix approximately the temperature at which the hot spots become visible. Experiments were carried out with a number of metal and metallic alloys, the melting-point of which covered a suitable range. The results are collected in table 4. The experiments showed that, when metals or alloys melting *below* 520°C were slid on glass or quartz, no hot spots could be seen even at the highest speeds and loads. With a gold aluminium alloy melting at 570°C , however, and with all metals melting *above* this, the hot spots were readily seen. This would fix the temperature at which the hot spots first become visible to the eye at between 520 and 570°C .

Photographic recording of hot spots

Although the occurrence of a transient hot spot is readily observable visually, the intensity is too low to affect a photographic plate. If, however, the slider is run over the same track a number of times, the cumulative effect is sufficient to produce a record on a photographic plate. A super XX plate ($6\frac{1}{2} \times 8\frac{1}{2}$ in.) was held in a frame mounted on a steel turn-table with the emulsion side upwards. A glass plate of the same size was clamped on top, its upper surface being ground to a uniform grease-free finish. The slider rested on the glass surface under a given load,

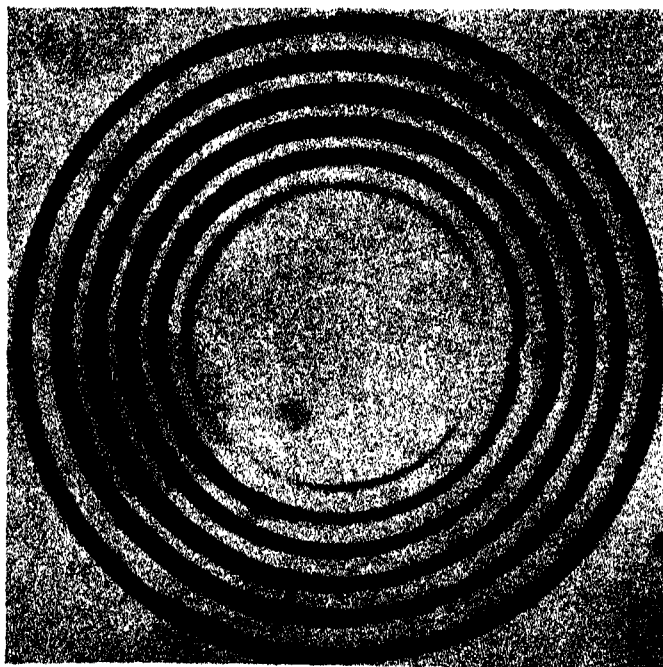


FIGURE 4. Photographic hot spot trace of steel slider on lapped glass surface. Load 1200 g. Speed of sliding at innermost visible track 70 cm./sec. approx.

and was allowed to run on the same track for 2 min. It was then moved in 1 cm. and again run for 2 min. The process was repeated and, in this way, a series of concentric tracks was obtained at various radii and therefore at various peripheral speeds. On developing the plate, a number of concentric dark rings appeared. The innermost visible ring on the plate gave the lowest speed at which hot spots could be recorded photographically under these conditions. A typical plate is shown in figure 4. It is seen that, with a load of 1200 g., a sliding speed of approximately 70 cm./sec. is just sufficient to produce a trace on the photographic plate. The results obtained for clean surfaces using four sliders of different thermal conductivities are shown in figure 5. In this figure the minimum speed to give a

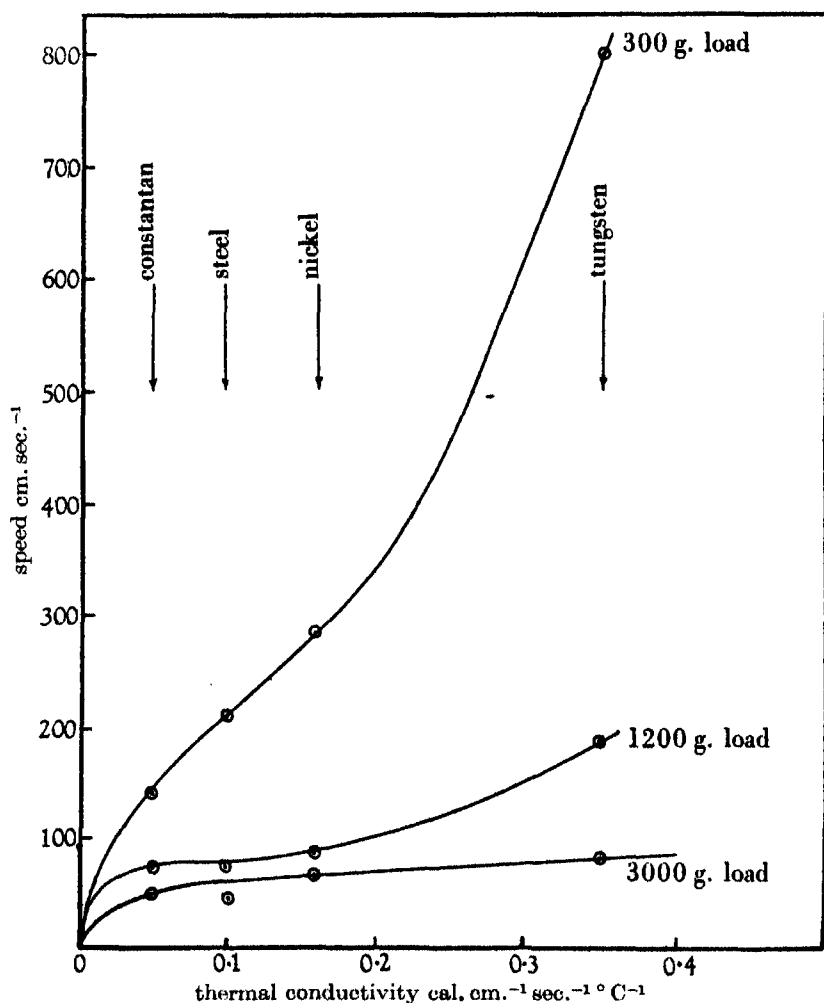


FIGURE 5. Conditions necessary to produce photographic records of hot spots for metal sliders running on a clean glass surface. The sliders used were constantan, steel, nickel and tungsten, with thermal conductivities of 0.05, 0.1, 0.16 and 0.35 cal. cm.⁻¹ sec.⁻¹ °C⁻¹ respectively.

record on the plate is plotted against the thermal conductivity for a number of fixed loads. The results are similar to those obtained by the visual method and the actual values of the loads and speeds agree with those obtained by visual observation. It was also found that the results obtained photographically for surfaces lubricated with glycerine and water were similar to those obtained visually.

The size and shape of the slider

The occurrence of hot spots is not influenced by the size or shape of the slider; it depends primarily upon the load and speed of sliding. If a large flat slider is used, the conditions of load and speed necessary are much the same as for a small curved slider. The main difference is that, with large flat surfaces, the hot spots may be thinly distributed over a wide area instead of being concentrated into a smaller one. This is in harmony with the view that contact between the solids occurs only locally at the summit of the surface irregularities, so that the real area of contact is very small and bears little relation to the apparent area of the surfaces. It means that even with light loads the *pressure* at the points of real contact is high and it is just at these points that the rubbing and the liberation of frictional heat occurs.

The effect of grit on the incidence of hot spots

A series of experiments was carried out to investigate the effect of grit on the incidence of hot spots. The experimental arrangement was the same as before. Carborundum of known particle size was used; this was mixed into a thick paste with glycerine and water and added to the glass disk. The disk was set rotating and the load increased until hot spots became visible.

It was found that with hard metal surfaces the presence of carborundum had very little effect. With metals such as constantan (V.P.H. 130) and tungsten (V.P.H. 750), the load and speed necessary to produce hot spots was approximately the same as that required with wet surfaces in the absence of abrasive, also the size of the particles (which was varied from 10 to 56 μ) made no appreciable difference to the incidence of hot spots.

With soft metals there are three general observations. First, when abrasive is present the glass disk is worn more rapidly over the region of the track. Secondly, after the surfaces have been running together, abrasive particles are always found embedded in the worn surface of the metal slider. The third observation is that, in the presence of abrasive, hot spots are observed very much more readily than when abrasive is absent. These results are shown fully in table 1. As a matter of interest, results are also included for an aloxite slider. This consists of a very hard cylinder of aluminium oxide abrasive, the particle size of the bonded grain being approximately 150 μ . The loads and speeds required to give visual hot spots with this slider on a wet surface are similar to those required by the soft metals in the presence of abrasive.

It is seen that, with grit, hot spots occur readily, even if the melting-point of the slider is well below 500° C. For example, with lead (m.p. 329° C) and tin (m.p. 230° C), hot spots occur at relatively low loads and speeds. No hot spots are observed with these metals in the absence of abrasive even at the highest loads and speeds. It should, however, be noted that, with a Wood's metal slider (m.p. 70° C, hardness at room temperature 16 Vickers), no hot spots were observed in the presence of abrasive and there was marked smearing of Wood's metal on the glass surface (cf. impact experiments, Bowden 1940).

TABLE 1. EFFECT OF GRIT ON THE INCIDENCE OF HOT SPOTS WITH SOFT METALS

metal	Vickers hardness	m.p. °C	thermal conductivity k cal. cm. ⁻¹ sec. ⁻¹ °C ⁻¹	load to give hot spots		abrasive paste (g.)	abrasive size μ	surface speeds cm./sec.	
				no abrasive clean (g.)	wet (g.)				
lead	4	327	0.08	↑ no hot spots observed at the highest loads and speeds ↓		400	56	780	
						1,200	56	300	
						1,200	28-40	300	
					tin	8	230	0.15	
	600	56	780						
	1,400	250	300						
	1,600	56	300						
zinc	80	418	0.26			1,200	28-40	300	
						2,100	14-20	300	
						400	14-20	780	
						1,200	56	300	
gold	26	1,060	0.7			1,200	14-20	300	
						5,500	0-10	300	
						700	14-20	780	
						2,500	56	300	
copper	60	1,080	0.9			2,800	14-20	300	
						6,000	0-10	300	
						10,000	0-10	190	
						1,000	14-20	780	
silver	55	960	0.98		275	—	5,000	56	300
					2,775	—	4,700	14-20	300
					2,775	—	8,000	0-10	300
					3,775	—	7,000	0-10	190
aloxite slider	—	—	2×10^{-3}		300	—	700	60	780
					300	—	1,200	14-20	780
					2,500	—	2,100	56	300
					2,500	—	2,100	28-40	300
aloxite slider	—	—	2×10^{-3}		2,500	—	5,000	14-20	300
					2,500	—	10,000	0-10	300
					100	200	200	14-20	780
					180	700	700	56	300
aloxite slider	—	—	2×10^{-3}		180	700	700	14-20	300
					700	2,100	2,400	14-20	110

It is apparent from table 1 that hot spots occur more readily the lower the thermal conductivity of the slider. The effect is thus similar to that observed in the absence of abrasive particles, although in the latter case the effect of thermal conductivity

is considerably more marked. This difference is shown by considering as an example the case of silver, which is a very good conductor ($k = 0.98$) and aloxite, which is a very bad conductor. For dry surfaces in the absence of abrasive when the peripheral speed is 300 cm./sec., the load required to produce visual hot spots with a silver slider is 2500 g., whilst with an aloxite slider it is 180 g.; the ratio is approximately 14:1. In the presence of abrasive of size $56\ \mu$ the values are 2100 and 700 g. respectively; the ratio in this case is about 3:1. With smaller abrasive particles, the thermal conductivity of the slider becomes more important.

The results given in table 1 show that, on the whole, hot spots occur more readily with the larger sized abrasive particles. Below a particle size of $20\ \mu$ the load necessary to produce visual hot spots increases very rapidly with decreasing particle size.

The experiments suggest that the occurrence of hot spots is due essentially to a layer of abrasive material embedded in the rubbing surface of the slider. On this view the slider acts as a lap, and the sliding really occurs between the embedded abrasive particles and the glass surface. This is shown by the increased wear of the glass disk, the appearance of the worn slider and the way in which the effect depends on the thermal conductivity of the slider and on the particle size of the abrasive. If the sliding surfaces are hard, the abrasive does not readily become embedded in the surface and it has a comparatively small effect.

We may make an approximate estimate of the temperature gradient set up in the abrasive particles during running. If we assume that visual hot spots correspond to a temperature of about 550°C and that the temperature of the metal slider cannot rise above its bulk melting-point, then, in the case of the lead slider, when hot spots are observed, there must be a temperature-drop of about 200°C through the abrasive particle, i.e. over a distance of about 10^{-3} cm . This implies a temperature gradient of the order of 10^5°C/cm . in the abrasive particle. These high gradients will, of course, be transient since the metal will rapidly soften around the particle allowing it to sink in, whilst the load will be borne by new abrasive particles coming into contact with the surface.

PART II. THE INITIATION OF THE EXPLOSION BY FRICTION

A series of experiments was carried out to determine the frictional conditions which would initiate the explosion of nitroglycerine present on the rubbing surface. The explosive was added to the freshly prepared surface, and the excess removed by running the disk for a few seconds. Under these conditions, only a small amount of explosive under, or in the immediate vicinity of, the slider was found to explode.

The influence of load, speed and shape of the slider

A known load was applied to the slider and the lower surface set in motion. When the load or speed reached a sufficiently high value, momentary explosions occurred

in the nitroglycerine. These explosive flashes varied very much in intensity, from a small crackling spark to a loud detonation. These effects were all grouped together, since, within the range of experimental observations, it is not possible to distinguish between the conditions necessary to produce a flash and those required for a loud explosion. It would seem that the intensity of the explosion was determined by the extent to which the initial explosion was *propagated* through the adjacent explosive and this was influenced by geometrical considerations and by the amount of explosive which had accumulated in the immediate vicinity of the slider. For this reason, the term 'explosion' is used loosely in the following section to describe the general phenomena of a small flash or a loud explosion.

Experiments were carried out on a variety of steel surfaces of widely different mechanical properties and on cast iron, and the speed and load necessary to produce explosions for different combinations of these are given in table 2.

TABLE 2. LOAD AND SPEED NECESSARY TO CAUSE EXPLOSION OF NITROGLYCERINE ON VARIOUS STEEL AND CAST IRON SURFACES

surfaces	sliding speed (cm./sec.)	load (kg.)	coefficient of friction	frictional force (F) (kg.)	FV (kg.cm./sec.)
mild steel on mild steel (turned)	55	7-9	—	—	—
	110	4-4.5	—	—	—
	220	2	—	—	—
	440	1	0.4	0.4	180
do. (lapped)	440	1	0.4	0.4	180
mild steel on chrome molybdenum	180	4.5	0.3	1.3	230
18 % tungsten steel on chrome molybdenum	90	5-6	0.4	2-2.5	180-270
	180	1.5-3	0.4	0.6-1.2	120-240
	360	1.5	0.3	4.5	210
carbon steel on chrome molybdenum	180	3-3.5	0.3	1-1.2	180-220
cast iron on mild steel (lapped)	440	1.5-2	0.3	0.5-0.6	210-260

It is interesting to compare the product FV (where F is the frictional force and V is the velocity of sliding) for the various combinations of sliding metals, since this is a measure of the rate of dissipation of frictional energy. The frictional force was measured by a spring device and, although in general it fluctuates violently, it was possible to obtain a fairly reproducible mean value for F . The results were similar for both finely lapped and for turned mild steel surfaces. With mild steel the wear was considerable. However, the results (i.e. the frictional force and speed necessary to cause explosion) were very similar to those obtained with a tungsten steel slider on a chrome molybdenum surface where the wear was very light. It will be seen that, over a wide range of load, speed, nature and hardness of the steel surfaces, the occurrence of explosions is dependent mainly upon the product FV .

(i.e. on the rate of dissipation of frictional energy). For example, when mild steel surfaces are used, the product of FV necessary to cause explosion of nitroglycerine is 180. The same value of FV obtains when a hard carbon steel slider is used on chrome molybdenum steel surface. These results are consistent with the view that, under these experimental conditions, explosion of the nitroglycerine is due to the frictional heating of the rubbing surfaces.

Experiments were also carried out with sliders of widely different geometrical shape. Hemispherical, conical, wedge shaped, cylindrical and flat sliders were used so that the apparent area of contact varied from point contact to that of comparatively large surfaces. It was found that the shape and size of the slider had no appreciable effect on the FV value at which explosion occurs. This is to be expected on the simple view (Bowden & Ridler 1936; Bowden & Tabor 1939) that the real area of contact, and hence the surface temperatures reached, is approximately independent of the apparent area of the surfaces. It was found, however, that if there was a small cavity present in one of the surfaces so that the explosive could accumulate in it, then the explosion, when it did occur, was much more violent.

Effect of thermal conductivity on the incidence of explosion

A series of experiments similar to those just described was carried out with the lower surface of glass covered with a thin film of nitroglycerine, and metal sliders of different thermal conductivity were used. As before, the plate revolved at a fixed speed and the load on the slider increased until explosion occurred. In some experiments the explosion occurred immediately; in others there was a time interval which varied from a period of a few seconds up to half a minute. In spite of this variation, the frictional force and speed at which explosion occurred were fairly well defined and reasonably reproducible. Figure 6 shows the results obtained. Each of the curves given in figure 6 is for a constant speed of revolution and shows a plot of minimum friction force to produce explosion against the thermal conductivity of the slider. These curves should be compared with those given in figures 2 and 3, which are plots of the slider conductivities against friction necessary to give visible *hot spots* on the surface.

It will be seen that the curves in figure 6 are very similar in form to those given in figures 2 and 3. It is also evident that the frictional force required to produce explosions is higher than that required to produce visible hot spots on clean surfaces (figure 2) and somewhat lower than that required to produce visible hot spots on wet surfaces (figure 3).

In table 3 the frictional forces necessary to produce hot spots on clean glass, or glass wet with glycerine and water mixture, are set out together with those necessary to produce explosions with nitroglycerine. Some results for a slider of fused silica are also included, and it is seen that the frictional forces involved are appreciably lower than those for constantan.

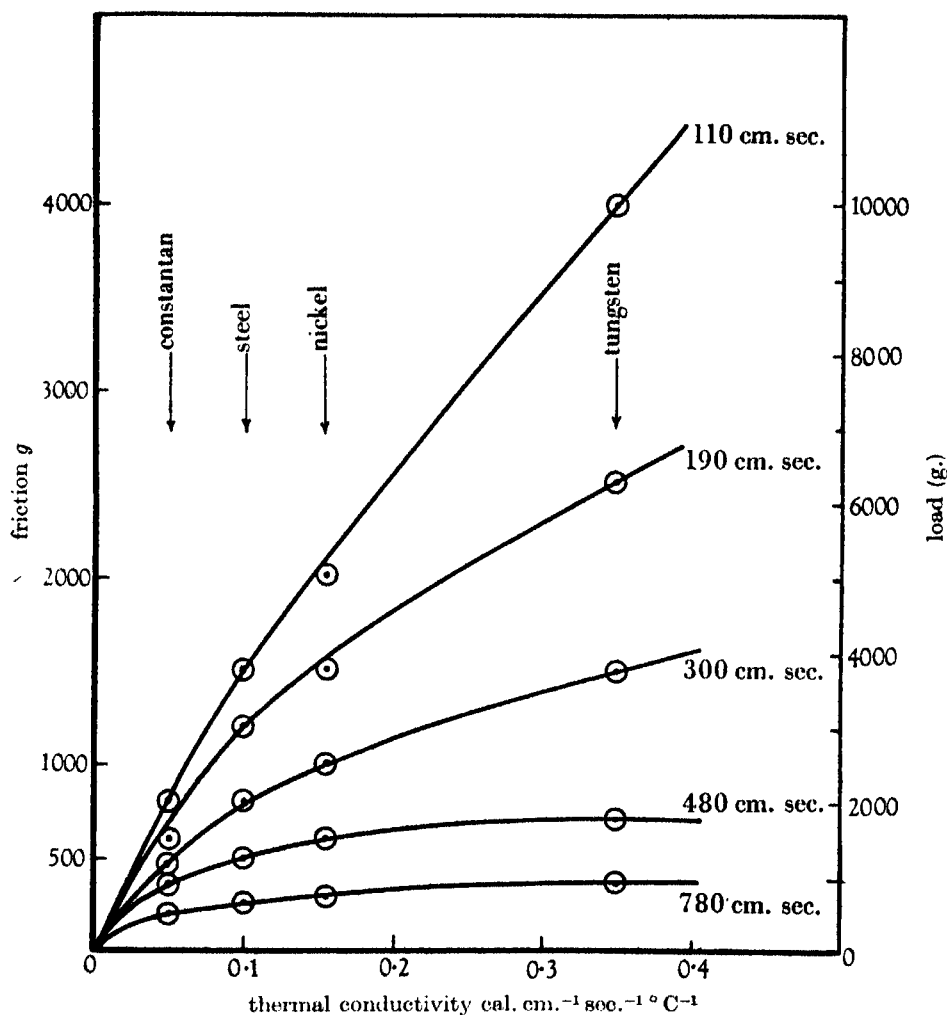


FIGURE 6. The explosion of nitroglycerine by friction for sliders of constantan ($k = 0.05$), steel ($k = 0.1$), nickel ($k = 0.16$) and tungsten ($k = 0.35$) rubbing on a glass surface. This figure should be compared with figures 2 and 3.

TABLE 3. EFFECT OF THERMAL CONDUCTIVITY ON INCIDENCE OF EXPLOSION

(lower surface glass: surface speed 300 cm./sec.)

slider	specific conductivity (cal. cm.⁻¹ sec.⁻¹ °C⁻¹)	frictional force required to give		
		hot spots on clean glass (g.)	hot spots on glass covered with glycerine and water (g.)	explosions with glass covered with nitroglycerine (g.)
silica (fused)	0.0035	120	180	140* 250
constantan	0.05	180	1400	500
steel	0.1	350	1800	800
nickel	0.16	400	—	1000
tungsten	0.35	500	4000	1500

* Nitroglycerine caught fire at this load and speed, but a higher load (250 g.) was needed to cause explosion.

Temperature of hot spots necessary to cause explosion

The use of a series of metal alloys of different melting-point to fix the temperature at which the hot spots first become visible has been described above. A similar set of experiments was carried out to determine the temperature necessary to produce explosion. A range of metals and alloys was used for the sliders, and these were selected so that they did not oxidize readily even at elevated temperatures. The lower surface was of glass and was covered with a layer of nitroglycerine, a slider of the appropriate metal was used and the load or the speed increased until explosion occurred. It was found that, with metals or metallic alloys melting below 450° C, no explosion occurred even at very high loads and speeds (see table 4). An alloy melting at 480° C and all metals melting above it, however, cause explosions. This indicates that the local high surface temperature which is just able to produce explosions under these conditions is about 480° C. This surface temperature is slightly lower than that necessary to give visible hot spots (520–570° C). The results of both series of experiments are collected in table 4.

TABLE 4. INCIDENCE OF HOT SPOTS AND EXPLOSIONS WITH METAL SLIDERS OF DIFFERENT MELTING POINTS ON GLASS

alloy	composition	m.p. °C	hard- ness V.P.H.	visual hot spots on clean glass	explosions of nitro- glycerine	comments
I	80 Au, 20 Sn	300	230	none	none	no hot spots or explosions even at highest loads and speeds
II	80 Au, 20 Pb	420	108	none	none	do.
III	75 Au, 25 Te	450	120	none	none	do.
IV	73 Ag, 27 Sn	480	93	none	several	no hot spots, but explosions at high loads
V	70 Ag, 30 Sb	480	120	none	several	do.
VI	80 Ag, 20 As	500	170	none	several	do.
VII	50 Au, 50 Cd	520	—	none	several	do.
VIII	92 Au, 8 Al	570	221	readily	readily	both hot spots and explosions
	constantan	1200	130	readily	readily	do.
	Ni	1450	170	readily	readily	do.
	Fe	1500	130	readily	readily	do.
	W	3000	—	—	—	do.

Effect of grit on the incidence of explosion

Experiments showed that, with soft metal sliders, the explosion occurred much more readily if particles of fine carborundum were present on the surfaces. It was also possible to obtain explosions with metals such as tin, which melt below 480° C if the abrasive were present. Grit therefore has a similar effect, both on the incidence of explosion and on the occurrence of visible hot spots.

PART III. THE THERMOELECTRIC MEASUREMENTS OF SURFACE
TEMPERATURE AND THE INCIDENCE OF EXPLOSION

In the final series of experiments a direct measurement was made of the surface temperature of the sliding surfaces, using the thermoelectric method, and a simultaneous record was made of the incidence of explosion. The lower surface was a rotating steel disk which was covered with nitroglycerine. The upper slider was of constantan and, during the actual run, it traversed from the outside of the disk to the centre so that a fresh surface was continuously presented to the slider. The flash from the explosion was picked up by a photomultiplier and reproduced with a strobotron on the moving film of the cathode-ray camera so that the incidence of the explosion was recorded simultaneously with the temperature trace.

Results

Portions of typical temperature traces are reproduced in figures 7-10, plates 10, 11. The strobotron flash which records an explosion may be seen as a vertical white line towards the bottom of the trace. As already mentioned, the strobotron is only capable of flashing approximately every $\frac{1}{150}$ sec. Explosions, however, may occur at much shorter intervals of time. To ensure that the flash of the strobotron corresponds to the instant of explosion, attention is concentrated only on the first of each group of flashes. In figures 8 and 9, plate 11, only the portions of the temperature trace before and up to a few thousandths of a second after the explosions are reproduced.

An examination of all temperature traces showed that the local surface temperature was fluctuating violently. Figure 7, plate 10 shows the results of a run taken with the camera started almost simultaneously with the main motor and before the sliding surfaces had attained full speed. For convenience, the trace has been reproduced in sections. Starting from the top, the trace progresses from left to right, the section directly underneath following on the right-hand end of the section above. The time marking is given by flashes of the small neon lamp occurring every 0.61 sec. and represented by the series of horizontal dashes directly under the temperature trace. It will be seen that, during the main part of the run, the speed of the camera was slowly increasing. The temperature, with the initial low sliding speed, was approximately 150° C. With the gradual increase in the speed of the plate, a rise in temperature was observed until towards the end of the first section a temperature of approximately 500° C was reached. At the beginning of the second section, the first strobotron flash indicates the first instant of explosion: the surface temperature was still about 500° C. The apparent drop in the temperature trace immediately following the explosion would appear to be due to the lifting effect of the explosion which separates the surfaces and causes an increase in contact resistance. The remaining strobotron flashes of the group are separated by a period approaching the minimum time interval of the strobotron and are neglected. The temperature trace is now seen to increase rapidly to its former

value and, with further increase in the speed of sliding, temperatures of the order of 700°C are reached towards the centre of the third section, where, evidently, the lower surface has attained its highest (constant) speed of rotation. A number of explosions occurs at this speed and temperature. From this point onwards the effect of continuously decreasing peripheral speed, due to the slider traversing from the outside of the plate towards the centre, becomes apparent in the gradual decrease in the average surface temperature and also in the number of explosions. Towards the end of the last section, the temperature has dropped to approximately 400°C and the last explosion occurred in the centre of the fourth section at a temperature of 600°C .

Typical temperature traces before and during an explosion taken at somewhat greater camera speeds, during two separate runs, are shown in figures 8 and 9, plate 11. In figure 8*a* and *b*, explosions have occurred at approximately 600°C . In the remaining two traces of figure 8, a somewhat higher temperature (700 – 750°C) existed at the instant of explosion. It will also be noted that explosions may not necessarily occur at the highest temperature of the trace. In figure 8*b*, a temperature of almost 800°C was reached only 0.004 sec. before explosion occurred at the considerably lower temperature of 600°C . This is also seen to a lesser degree in figure 8*c*. In figure 9*a* and *b*, the temperature at the explosion has reached a value of approximately 700°C , while in 9*c* and *d* the temperature is of the order of 600°C .

It is found that explosion does not necessarily take place at the highest temperature. This is illustrated by two traces shown in figure 7, plate 10 taken at the maximum camera speed. In the top trace, an explosion has taken place at approximately 500°C although the temperature at this point is increasing and the peak temperature in the trace is consistently 600°C . A similar effect, not quite so well defined, is observed in the lower trace. In this case the temperature at which the explosion occurred was approximately 500°C .

DISCUSSION

It will be seen that visible hot spots occur very readily on the surface of non-conductors when they are slid over one another. When one of the surfaces is of glass or quartz and the other of metal (e.g. constantan), hot spots may be seen when the sliding speed is as low as one or two feet per second and the load is about 1000 g. If the upper slider of metal is replaced by a non-conductor such as quartz, the hot spots appear even more readily. The experiments with metallic sliders of different melting-points show that, under the conditions of these experiments, the hot spots first become visible when the surface temperature is about 520 – 570°C . As we should expect, the load and speed at which the hot spots appear is markedly dependent upon the thermal conductivity of the slider but is little influenced by the size of the rubbing surfaces. Grit, in the form of small particles of carborundum, has little influence on hard metals, but on soft metals, in which it becomes embedded, it causes hot spots to appear more readily. With any given combination of

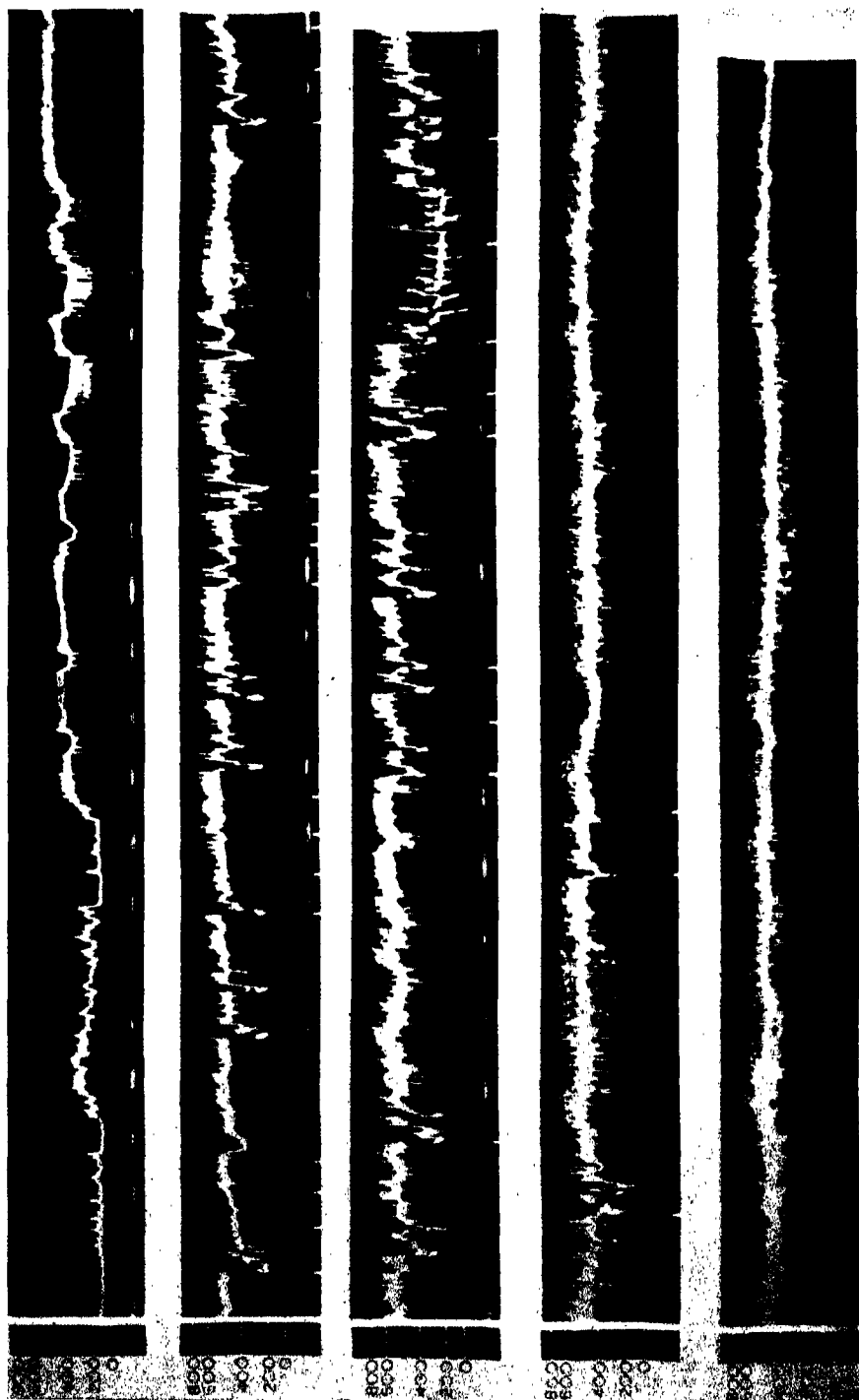


FIGURE 7. The first of the group of vertical white marks at the bottom of each photograph indicates the explosion. Constantan slider on lapped steel surface. Load 3700 g.

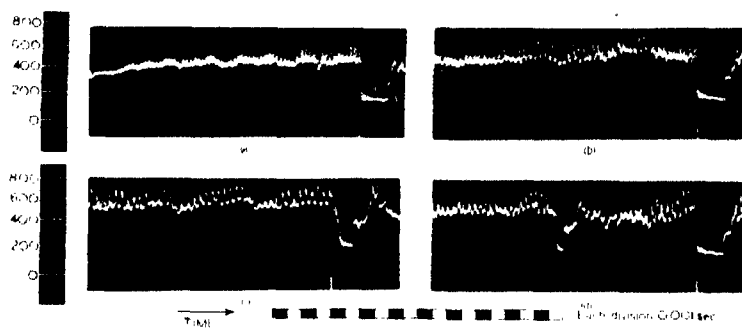


FIGURE 8

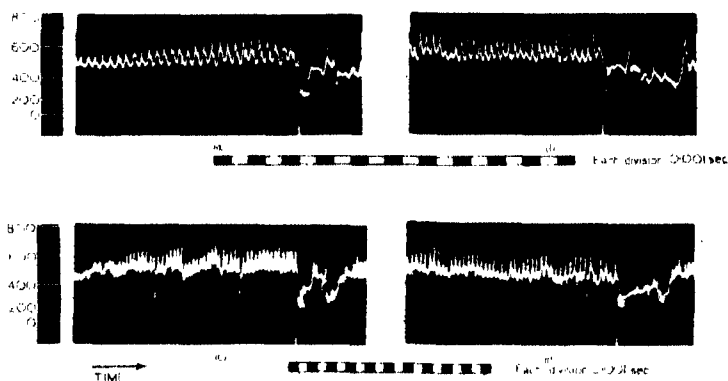


FIGURE 9

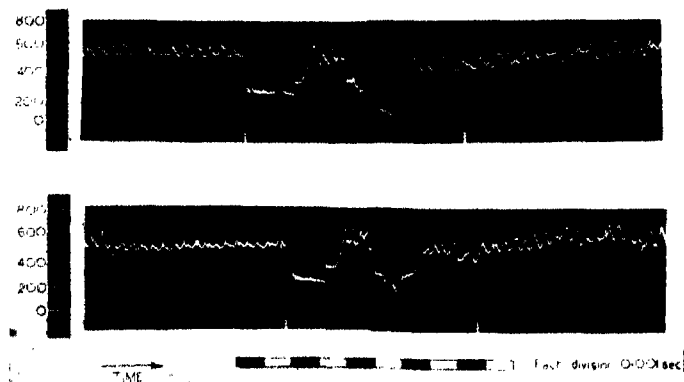


FIGURE 10

FIGURES 8-10. The *first* vertical white mark at the bottom of each figure indicates the explosion.

surfaces the occurrence of hot spots depends primarily upon the rate of dissipation of frictional energy, that is, it depends mainly on the product FV , where F is the frictional force and V the velocity of sliding. If the surfaces are wetted with glycerine and water the hot spots still appear but the frictional force necessary to produce them is greater by a factor of six or seven.

If the surfaces are wetted with a liquid explosive such as nitroglycerine, the incidence of explosion with any given combination of surfaces is again determined mainly by the product FV . When sliders of different metals are used, the load or the speed necessary to cause explosion increases regularly with the thermal conductivity of the metal, and the results run parallel to those obtained in the recording of hot spots. The experiments with metals of different melting-points fix the surface temperature at which explosion first occurs as 480°C .

When both the sliding surfaces are of metal and the surface temperature is measured by the thermoelectric method, the results are similar. Since both surfaces are now good thermal conductors, the loads and speeds required to cause explosion of the nitroglycerine are higher. The simultaneous measurement of the surface temperature and of the incidence of the explosion show that explosion occurs when the local surface temperature of the rubbing metals reaches about 500°C . In all these experiments no explosion was observed when the recorded surface temperature was appreciably below this. It is also evident that the explosion does not necessarily take place as soon as this temperature is reached. A temperature as high as 800°C frequently occurs for short intervals of time without explosion of the film. It would therefore seem that although the occurrence of a momentary local high temperature of about 500°C is a necessary condition for explosion under these conditions, it may not in itself be enough. We should expect that other factors such as the presence of sufficient explosive at the hot spot and the size, number and duration of the hot spots themselves would all be important. Unless these conditions were suitable both for the initiation of the explosive reaction and for its propagation to an appreciable distance, no explosion would be observed. If the conditions are not suitable, the moving hot spots may be quenched by the surrounding cold nitroglycerine or by moving on to a cold area of the surface before an explosion of finite size can take place.

The size of the hot spots on the rubbing surface is not known, but our knowledge of the shape and size of the surface irregularities and the appearance of the visible hot spots suggests that their size is variable and that some of them may be of the order of 10^{-3} cm. in diameter. It is well known from the work of Robertson (1909), Roginsky (1930) and others that nitroglycerine undergoes thermal decomposition at temperatures well below that at which explosion occurs, and the decomposition obeys a unimolecular law between 90 – 190°C . At 190°C the period of half-life is several seconds. If nitroglycerine is heated suddenly to temperatures of about 200°C , explosive decomposition occurs after an induction period. Unfortunately there is little information available about the induction period of nitroglycerine, but at 200°C it is appreciable. This would readily explain why hot

spots of 200°C or so on the moving surfaces are ineffective in causing explosion. In the friction experiments the sliding speed was usually about 100 cm./sec. or more, so that in one-hundredth of a second the hot spots will have moved forward 1 cm. They would therefore be quenched before appreciable decomposition could occur. Both the induction period of nitroglycerine and the period of half-life decreases rapidly as the temperature is raised, but their values at high temperatures are not known. If it is assumed that the unimolecular relation holds up to temperatures of 400°C , the period of half-life at this temperature would be about 10^{-6} sec. The impact experiments on the initiation of nitroglycerine by the adiabatic compression of small gas bubbles (Bowden, Mulcahy, Vines & Yoffe 1946) show that the delay between compression and initiation is very short and may be only a few microseconds. If we take a value of say 10^{-5} sec., the hot surface will in this time have only moved forward a distance of 10^{-3} cm. so that initiation may well occur before the hot spot is quenched. A detailed analysis of the processes occurring at a hot spot would be valuable and further work on this is in progress. There is no indication from figure 7, plate 10 and figure 8, plate 11, that the explosion contributes materially to the temperature of the hot spots, and this suggests that there is probably a short time lag before explosion occurs, similar to that observed with initiation by small bubbles.

In these experiments the explosive is subjected to heavy local pressure and to the shearing action of the rubbing solids, but apparently this in itself is not able to cause initiation. When steel surfaces are used, for example, they are deformed and torn during the sliding process. This means that the local pressures between the surfaces has exceeded the flow pressure of steel. In the case of mild steel the hardness was approximately 150, and with chrome molybdenum surfaces 300 kg./mm.². In the latter case this would correspond to a pressure of 30,000 atm. It is clear from the results given in table 4 that the occurrence of the explosion is not determined by the hardness of the sliding surfaces. Tungsten, for example, which is very hard, is much less effective than the softer constantan because its thermal conductivity is higher. Similarly it will be seen from table 3 that the occurrence of the explosion is controlled by the melting-point of the metal and not by the hardness; a hard gold-tin alloy (V.P.H. 230) which melts at 230°C will not produce an explosion whereas a much softer silver-tin alloy (V.P.H. 93) which melts at 480°C will cause explosion.

If the metal is a good thermal conductor, sliding speeds of 700 cm./sec. can be used without explosion. Since the layer of explosive between the rubbing surfaces is necessarily very thin, this would correspond to a high rate of shear. The conditions are those of boundary lubrication since certain areas of the surface are separated by a few molecular layers at most. If we assume that a monolayer of explosive adheres to each surface and that shear occurs through a few molecular layers of nitroglycerine, the velocity gradient may be of the order of 10^9 sec.⁻¹. These high pressures and high rates of shear, however, do not produce an explosion unless the local surface temperature rises to about 500°C .

These results illustrate how difficult it is to bring about the decomposition of nitroglycerine by direct mechanical means. Although the nitroglycerine is subjected to violent rubbing and grinding at very heavy loads and high speeds, it fails to explode unless the temperature of the surfaces is raised to a sufficiently high value. Under the conditions of these experiments it is apparently necessary that the mechanical energy should first be degraded into heat. The local hot spots on the surfaces then bring about the decomposition of the explosive.

We wish to express our thanks to Dr D. Tabor for valuable discussions and assistance throughout the progress of this work, and to Mr A. E. Ferguson, B.E.E., and Mr R. W. Muncey, B.E.E., for the design of the electrical apparatus.

REFERENCES

- Bernal, J. D. 1938 *Trans. Faraday Soc.* **34**, 1008.
Bowden, F. P. 1940 Unpublished Report.
Bowden, F. P. & Hughes, T. P. 1937 *Proc. Roy. Soc. A*, **160**, 575.
Bowden, F. P., Mulcahy, M. F. R., Vines, R. G. & Yoffe, A. 1947 *Proc. Roy. Soc. A*, **188**, 311.
Bowden, F. P. & Ridler, K. E. W. 1936 *Proc. Roy. Soc. A*, **154**, 640.
Bowden, F. P. & Tabor, D. 1939 *Proc. Roy. Soc. A*, **169**, 391.
Robertson, R. 1909 *J. Chem. Soc.* **95**, 1241.
Roginsky, S. 1930 *Phys. Z. Sowjet.* **1**, 640.
Taylor, W. & Weale, A. 1938 *Trans. Faraday Soc.* **34**, 995.

The quantal calculation of the photo-ionization cross-section of atomic potassium

By D. R. BATES, M.Sc., *University College, London*

(Communicated by H. S. W. Massey, F.R.S.—Received 20 February 1946)

The photo-ionization of atomic potassium is investigated using quantal methods. Emphasis is laid on the great sensitivity of the cross-section to the wave functions employed. The general features of the cross-section as revealed by the laboratory measurement can be understood. To explain the finite minimum observed in the cross-section-frequency curve it is necessary to take into account non-separability effects.

1. INTRODUCTION

Measurements of the photo-ionization cross-section of atomic potassium have been made by Ditchburn, Tunstead & Yates (1943). In confirmation of the early results of Ditchburn (1928), Lawrence & Edlefsen (1929) and others,† it was found that as the frequency of the incident radiation is increased above the spectral head, the cross-section falls, reaches a minimum and then rises rapidly. This unusual effect is at variance with the quantal calculations of Phillips (1932) which had predicted merely a steady decrease. In addition, the apparent agreement between experiment and theory in giving the magnitude of the cross-section to be low (of order 10^{-20} cm.² near the spectral head) is fortuitous, as in the latter a multiplying constant appears to have been taken to be much too small.

In a recent study of the photo-ionization of calcium Bates & Massey (1941) showed that electron exchange (which, of course, was not taken into account in the work of Phillips) could profoundly alter the cross-section. This clearly made desirable a reconsideration of the potassium problem to see if the lack of agreement with observation could be removed. It was decided to repeat the calculations using very accurate wave functions in which allowance was to be made not only for electron exchange, but also for distortion and polarization of the core. Before this was completed, however, it became manifest that, owing to the extremely high degree of interference in the matrix element involved in the transition, the final reliability would not be sufficient to justify the considerable computational labour necessary.‡ The investigation, therefore, was directed rather towards determining if the experimental results could be reproduced by using wave functions that appeared acceptable.

† See review by Page (1939).

‡ It may be pointed out, however, that the sensitivity of the cross-section would be useful if it were desired to check the accuracy of the wave functions given by any particular model.

2. THE METHOD OF THE CALCULATIONS

2.1. The formula for the photo-ionization cross-section k_ν of an atom to radiation of frequency ν is simply

$$k_\nu = \frac{32\pi^4 m^2 e^2}{3\hbar^3 c} \nu v a \left| \int \chi_i(\Sigma \mathbf{r}) \chi_f^* d\tau \right|^2, \quad (1)$$

where v is the velocity of the ejected electron, a is a weighting factor, χ_i is the initial wave function of the atom normalized to unit density, \mathbf{r} is the displacement of a particular electron from the nucleus, and χ_f is the final wave function of the positive-ion electron complex normalized to represent a core function of unit density together with a wave of unit amplitude at infinity. In the case of an atom like potassium, the outer electron only need be taken into account in the transition. Moreover, as in the bound state the azimuthal quantal number l of this is zero, in the free state it must be unity. Reduction in formula (1) can at once be effected to give

$$k_\nu = \frac{\hbar^5}{6\pi^3 m^3 e^4 c} \frac{\nu}{v} \left| \int_0^\infty P(4s|r) r P(\epsilon p|r) dr \right|^2, \quad (2)^\dagger$$

where $P(4s|r)/r$ is the radial wave function of the bound state normalized so that

$$\int_0^\infty P(4s|r)^2 dr = 1, \quad (3)$$

and $P(\epsilon p|r)/r$ is the radial wave function of the free state normalized so that, at large r ,

$$P(\epsilon p|r) \rightarrow \left(\frac{\pi e^{\frac{1}{2}}}{2} (1+c) \right)^{\frac{1}{2}} \frac{a G_1(\epsilon|r) + b H_1(\epsilon|r)}{(a^2 + b^2)^{\frac{1}{2}}}, \quad (4)$$

$G_1(\epsilon|r)$ and $H_1(\epsilon|r)$ being functions introduced by Hartree (1928) and Hargraves (1929) and tabulated by Bates & Massey (1941) and a and b being constants. Substituting numerical values and taking the ionization potential of potassium as 4.32 eV, (2) becomes

$$k_\nu = 2.71 \times 10^{-19} \left\{ \frac{1 + 3.13\epsilon}{\epsilon^{\frac{1}{2}}} \right\} \left| \int_0^\infty P(4s|r) r P(\epsilon p|r) dr \right|^2 \text{ cm.}^2. \quad (5)$$

In order to complete the calculation $P(4s|r)$ and $P(\epsilon p|r)$ have to be found and the matrix element computed.

2.2. A table of $P(4s|r)$ has been published by Hartree & Hartree (1938*b*). It was obtained by the Fock method and includes the effect of electron exchange. Though no allowance was made for core distortion (in the mean field of the active electron) and polarization (in the instantaneous field of the active electron) so that

† The quantities in the external factor are expressed in c.g.s. and those under the integral sign in Hartree units (radial distance r , in units a_0 , the radius of the first Bohr orbit of hydrogen; and the kinetic energy of the ejected electron ϵ , in units I_H , the ionization potential of hydrogen).

the function must tend to be rather too diffuse, it is sufficiently accurate to enable an estimation to be made of the theoretically possible cross-section curves. It had, however, to be extrapolated to greater radial distances than those given.

2.3. Ideally in calculating $P(\epsilon p | r)$ the atomic field should be corrected for core polarization† and the Fock exchange equation should be set up and solved. Instead of this elaborate procedure it was decided to neglect electron exchange and to treat the polarizability of the core as a parameter to be varied over sufficiently wide limits to cover the uncertainty in the exact form of the wave functions. Using Hartree units (1928), the differential equation satisfied by $P(\epsilon p | r)$ is thus

$$(d^2/dr^2 + (2Z(r)/r - 2V_P(r) - 2/r^2 + \epsilon)) P(\epsilon p | r) = 0. \quad (6)$$

Here $Z(r)$ is the effective nuclear charge of the potassium positive ion. It has been determined by the Hartrees (1938*a*). $V_P(r)$ is the additional potential due to polarization and was taken to be given by

$$V_P(r) = -\frac{1}{2}(P/(r^2 + \rho^2)^2), \quad (7)$$

where P is the polarizability and ρ the radius of the core. The experimental polarizability is approximately 0.9×10^{-24} cm.³, but the effects of different values from zero to 2×10^{-24} cm.³ are of interest. It should be noted that polarizabilities higher than the experimental tend to compensate for the neglect of electron exchange as this is attractive for potassium (cf. the positive ion of calcium—Bates & Massey 1941). ρ , which is relatively unimportant, was taken as $1.4a_0$. By means of a power series near the origin and numerical integration at greater distances, solutions to equation (6) for $P(\epsilon p | r)$ were found for the range of polarizabilities just mentioned and for kinetic energies of the ejected electron up to $0.1I_H$ (1.35 eV). These were normalized by the relation (4).

2.4. Finally, the various wave functions were employed in the evaluation of the matrix elements $\left| \int_0^\infty P(4s | r) r P(\epsilon p | r) dr \right|$ from which the photo-ionization cross-sections were immediately obtained by formula (5).

Since there is no doubt as to the validity of the fundamental formula, and since the basic wave functions are certainly sufficiently accurate to justify the use of the variable parameter, the main features of the cross-section-energy curve found in practice should be included amongst the results.

3. THE RESULTS

3.1. The calculated photo-ionization cross-sections for atomic potassium with various assumed polarizabilities are given in table 1. At the spectral head the cross-sections are comparatively large for low polarizabilities but become smaller as

† Core distortion can be ignored at the experimental densities of free electrons.

the polarizability is increased reaching a zero minimum for a polarizability of 1.6×10^{-24} cm.³. Near the minimum they are extremely sensitive to the exact magnitude of the polarizability. The polarizability has also a great influence on the shape of the curve of cross-section against energy. In the region involved the cross-section falls off steadily from the value at the spectral head for low polarizations, but for higher polarizabilities it develops a zero minimum at some energy above. As the polarizability is further increased the location of this minimum moves in towards and eventually attains the spectral head. For still higher polarizabilities the cross-section rises with energy tending to a broad maximum.

TABLE 1. PHOTO-IONIZATION CROSS-SECTION FOR ATOMIC POTASSIUM

kinetic energy of ejected electron in units I_H (13.5 eV)	polarizability (cm. ³)				
	0.0×10^{-24}	0.5×10^{-24}	1.0×10^{-24}	1.5×10^{-24}	2.0×10^{-24}
	cross-section (cm. ²)				
0.000	6.4×10^{-18}	2.8×10^{-18}	7.6×10^{-19}	2.2×10^{-20}	2.5×10^{-19}
0.025	4.9	2.0	4.9	0.2	3.0
0.050	3.8	1.5	3.2	0.1	3.3
0.075	3.1	1.2	2.1	0.8	3.6
0.100	2.5	0.9	1.4	2.2	3.7

3.2. The reasons for the effects described can best be understood from figures 1 and 2, which show representative plots of $\left(\int_0^\infty P(4s|r)rP(ep|r)dr\right)$ against P , and $(P(4s|r)rP(ep|r))$ against r respectively. As can be seen in figure 1 the value of the integral decreases uniformly with increasing polarizability, passes through zero and becomes negative. This behaviour arises from the tendency towards cancellation of the positive and negative portions of the integrand, the former being the smaller at low polarizations and the latter at high polarizations (figure 2). The extent of this cancellation is of the utmost importance in estimating the reliance to be placed on any results obtained. A convenient measure of it is got by multiplying the integral by a suitable factor so that it can be written in the form $(1-D)$, where D is lesser of the two contributions of opposite sign. For some different cross-sections the values found for $(1-D)$ at the spectral head are given below:

cross-section (cm.) ²	5×10^{-18}	5×10^{-19}	1×10^{-19}	2×10^{-20}	1×10^{-20}	5×10^{-21}
$(1-D)$	0.388	0.143	0.067†	0.030†	0.021†	0.015†

The corresponding values at energies above the spectral head are slightly greater. Bearing in mind that the cross-section depends on $(1-D)^2$, the high degree of accuracy necessary in both the initial and the final wave functions is clear. It must be emphasized that even at large radial distances considerable accuracy has to be maintained (especially near the spectral head). This is illustrated by the magnitude

† These figures are the mean of the slightly different results obtained from the two possible polarizabilities.

of the cross-sections obtained if a $\pm 10\%$ error is made in $(P(4s|r)rP(\epsilon p|r))$ for value of r greater than $16a_0$. Thus taking the same examples as above

true cross-section (cm.) ²					
5×10^{-18}	5×10^{-19}	1×10^{-19}	2×10^{-20}	1×10^{-20}	5×10^{-21}
false cross-sections (cm.) ²					
5.1×10^{-18}	5.5×10^{-19}	$1.3 \times 10^{-19} \dagger$	$3.4 \times 10^{-20} \dagger$	$2.0 \times 10^{-20} \dagger$	$12.7 \times 10^{-21} \dagger$
4.9×10^{-18}	4.5×10^{-19}	$0.8 \times 10^{-19} \dagger$	$1.0 \times 10^{-20} \dagger$	$0.3 \times 10^{-20} \dagger$	$0.8 \times 10^{-21} \dagger$

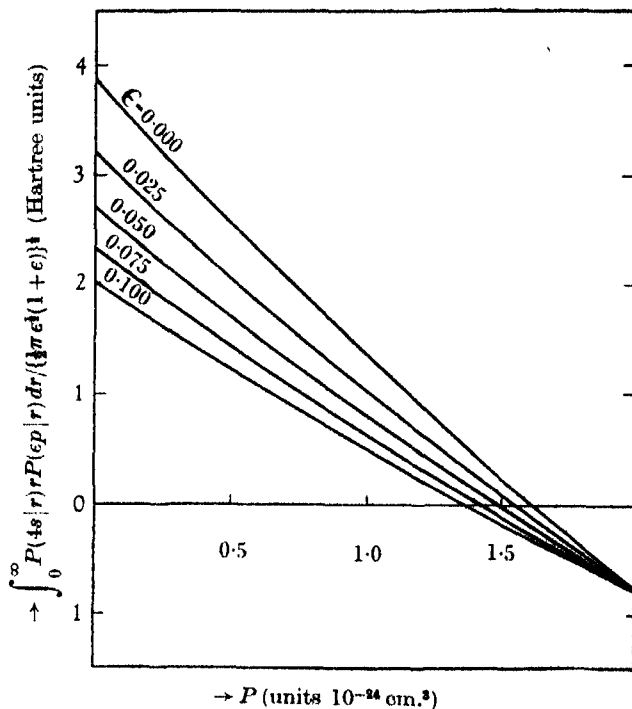


FIGURE 1. The variation of the transition integral with polarizability for various energies of the ejected electron. ϵ in units I_H .

4. DISCUSSION

As stated earlier the photo-ionization cross-section measured by Ditchburn *et al.* is only of order 10^{-20} cm.² near the spectral head. Even if the view be taken that the experimental results arose from some type of absorption not allowed for in the theory, the very significant fact remains that *the simple process actually considered cannot have a greater cross-section*. In consequence the polarizability to be used *must* be close to the value 1.6×10^{-24} cm.³.[†] Figure 3 shows cross-section wave-length curves

[†] These figures are the mean of the slightly different results obtained from the two possible polarizabilities.

[‡] Though such a conclusion would not have been reached without experimental aid it at least is not at variance with expectation.

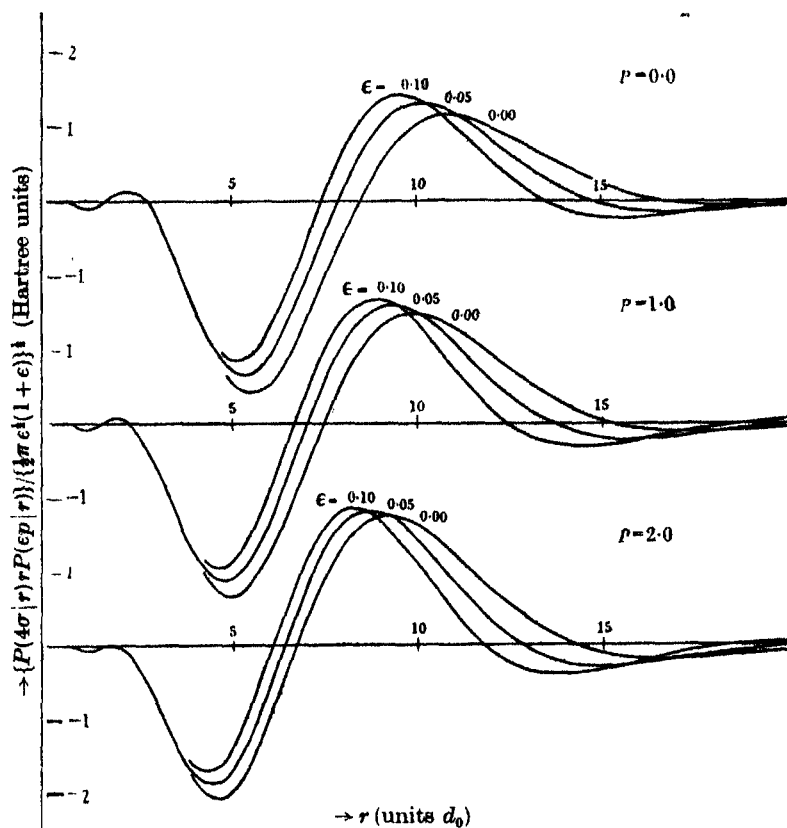


FIGURE 2. The form of the integrand for various polarizabilities and various energies of the ejected electron. P in units 10^{-24} cm.³, ϵ in units I_H .

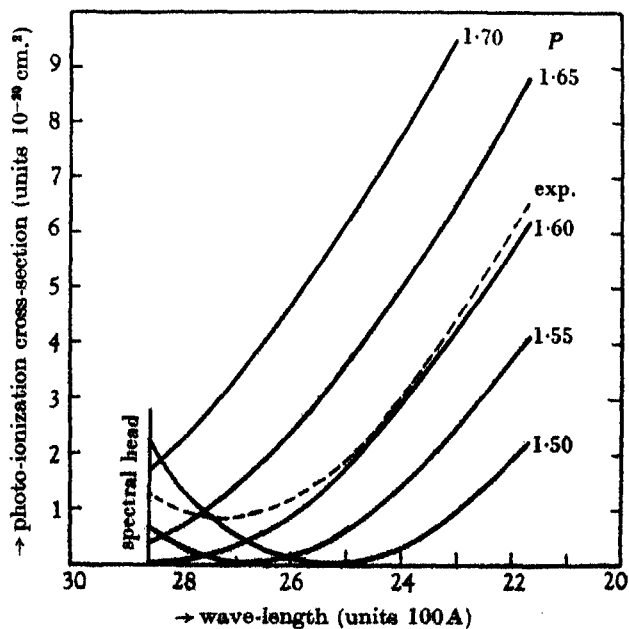


FIGURE 3. Cross-section wave-length curves. Full lines: theoretical curves for various polarizabilities. Dashed line: experimental curve. P in units 10^{-24} cm.³

calculated with polarizabilities of 1.50, 1.55, 1.60, 1.65 and 1.70 10^{-24} cm.³ (and for comparison the experimental curve). It can be seen that meeting the requirement for low magnitude of the cross-section leads necessarily to unusual variations with wave-length not dissimilar in general features to those observed. Detailed agreement is not obtained, but in view of the dependence on the exact form of the wave functions this is not serious except in that the theory does not give a finite minimum. The reason for the occurrence of this in practice is not at first obvious.

(i) Any effect from collisions (even if appreciable) must be pressure dependent and thus is eliminated by the method of analysis of the measurements adopted to separate the absorption by atomic and by molecular potassium.

(ii) Both quadrupole transitions and the consequence of the departure from perfect spherical symmetry of the potential in which the active electron moves were considered but were found to be inadequate to account for the discrepancy.

(iii) It was assumed in the investigations that complete separability of the wave functions involved is possible. Thus they were written simply as the product of the wave functions of the ground state of the core and that of the electron in the normal field. However, the complete solutions of the wave equation contain in addition contributions from the other members of the orthogonal set of core wave functions and can in fact be considered as expansions in the form

$$\chi_i(\mathbf{r}_c, \mathbf{r}) = \chi(X^0 | \mathbf{r}_c) \chi_i(4s | \mathbf{r}) + \sum_{X', l} \chi(X' | \mathbf{r}_c) \chi_l(El | \mathbf{r})^\dagger \quad (8)$$

and
$$\chi_f(\mathbf{r}_c, \mathbf{r}) = \chi(X^0 | \mathbf{r}_c) \left\{ \sum_l \chi_l(el | \mathbf{r}) \right\} + \sum_X \chi(X' | \mathbf{r}_c) \left\{ \sum_l \chi_l(El | \mathbf{r}) \right\}, \quad (9)$$

where $\chi(X^0 | \mathbf{r}_c)$ is the ground core wave function, and $\chi(X' | \mathbf{r}_c)$ is an excited core wave function. For the system involved the energy E of the active electron is of course negative when associated with an excited core, but the amplitude of the relevant wave functions $\chi_l(El | \mathbf{r})$, $\sum \chi_f(El | \mathbf{r})$ may remain appreciable. The transition probability can be calculated as usual by applying formula (1) to the wave functions $\chi_i(\mathbf{r}_c, \mathbf{r})$, $\chi_f(\mathbf{r}_c, \mathbf{r})$. Combinations between certain of the excited core terms and the ground core terms can clearly occur. Owing to the introduction of different complex multipliers by the various phases of the free orbitals the modulus of the total matrix element does not pass through zero. Unfortunately, the wave functions of neither the excited core nor of the corresponding active electron have been determined. The equations for the former are of the standard type. Those for the latter have been discussed by Massey & Bates (1940) and need not be given here. The numerical work necessary for an accurate evaluation is very extensive and is scarcely warranted by the degree of interest attached to the magnitude of the cross-section at the minimum. In consequence it was merely verified by crude approximations that the correction terms can give rise to an effect of the right order.

† Many of the terms here are of course absent—all those for example of odd parity.

5. CONCLUSION

If allowance is made for the non-separability of the wave functions it seems probable that the experimental results can be completely explained, the close cancellation of the major part of the matrix element giving rise to the abnormally small cross-section and the anomalous frequency variation.

Acknowledgement and thanks are due to Professor Massey for many helpful discussions.

REFERENCES

- Bates, D. R. & Massey, H. S. W. 1941 *Proc. Roy. Soc. A*, **177**, 329.†
Ditchburn, R. W. 1928 *Proc. Roy. Soc. A*, **117**, 486.
Ditchburn, R. W., Tunstead, J. & Yates, J. G. 1943 *Proc. Roy. Soc. A*, **181**, 386.
Hargraves, J. 1929 *Proc. Camb. Phil. Soc.* **25**, 75.
Hartree, D. R. 1928 *Proc. Camb. Phil. Soc.* **24**, 89, 426.
Hartree, D. R. & Hartree, W. 1938a *Proc. Roy. Soc. A*, **166**, 450.
Hartree, D. R. & Hartree, W. 1938b *Proc. Camb. Phil. Soc.* **34**, 550.
Lawrence, E. O. & Edlefsen, N. E. 1929 *Phys. Rev.* **34**, 1056.
Massey, H. S. W. & Bates, D. R. 1940 *Astrophys. J.* **91**, 202.
Page, T. L. 1939 *Mon. Not. R. Astr. Soc.* **99**, 385.
Phillips, M. 1932 *Phys. Rev.* **39**, 905.

† This opportunity is taken of correcting an error that appeared in this paper. In deriving the electronic recombination coefficient to Ca^+ from the photo-ionization cross-section of Ca an incorrect weighting factor was used. As a result the value given is too small by a factor of 2.

Structure and thermal properties associated with some hydrogen bonds in crystals

VII. Behaviour of KH_2PO_4 and KH_2AsO_4 on cooling

BY A. R. UBBELOHDE AND I. WOODWARD

(Communicated by Sir Henry Dale, F.R.S.—Received 22 February 1946)

[Plate 12]

A low-temperature micro-thermostat is described, suitable for single-crystal X-ray work. Thermal expansions of KH_2PO_4 (and in less detail of KH_2AsO_4) have been measured by X-ray methods from room temperatures down to about 80°K .

Both these crystals differ from some previously investigated organic compounds with short hydrogen bonds, in that maximum thermal contraction on cooling occurs along the c axis, which is approximately perpendicular to the planes containing the hydrogen bonds in these two crystals. This axis becomes electrically polarized below a thermal 'transition point'.

Remarkable effects are observed in the neighbourhood of the transition point, similar to those previously described for Rochelle salt. When the tetragonal single crystals are cooled below the respective Curie points (123°K for KH_2PO_4 and 96°K for KH_2AsO_4) they change into crystal 'hybrids', in which subcrystalline units of lower symmetry pack so as to preserve the c axis and either the (100) or (010) planes of the original tetragonal structure, and to maintain electrical neutrality on the average. When a 'hybrid' crystal is warmed above the Curie point, the subcrystalline units merge after a varying lapse of time into the original tetragonal lattice. This change from single crystal to hybrid and back can be repeated at will by cooling and heating. The structure of these crystal hybrids has been investigated by both Bragg and Laue methods.

On the basis of the new experimental evidence and of previous work, an explanation is offered for phase transitions 'of the second kind', preferably termed 'continuous', in terms of crystal structure. It is known that 'discontinuous' phase changes at a transition 'point' involve a rearrangement of molecules within the single crystal of one structure, to give a powder of crystals with a different structure. But if the new crystals can pack as subcrystalline units into the original crystal lattice, as is possible when the difference in structure is small and not too much strain is involved, the phase change will occur continuously over a range of temperatures instead of discontinuously at a transition point.

The X-ray studies described give new information about the thermodynamics of crystals, and promise to throw light on certain problems of imperfect crystals. Possible biological analogues of hybrid crystals are briefly referred to.

INTRODUCTION

It has been previously established that comparatively small changes in bond lengths, such as result on replacing H by D, are sufficient to upset the crystallization of KH_2PO_4 in the tetragonal form (Ubbelohde 1939*b*). Interest in thermal effects in KH_2PO_4 was heightened by recent X-ray observations on Rochelle salt (Ubbelohde & Woodward 1946), which established that below the upper Curie point single crystals changed into 'hybrids' composed of subcrystalline monoclinic units oriented so as to preserve the a axis of the original crystal as common symmetry

axis. Since dielectric anomalies similar to the Rochelle salt effect had recently been reported in KH_2PO_4 and KH_2AsO_4 (Busch 1938), it was natural to look for a corresponding formation of hybrid crystals, using our X-ray methods.

A rapid survey using drips of liquid oxygen or liquid nitrogen on to crystals of these salts gave evidence of hybrid crystal formation at low temperatures in a manner similar to Rochelle salt but even more beautifully defined (Ubbelohde & Woodward 1945). More detailed experiments were therefore carried out:

(1) To devise a simple, low-temperature micro-thermostat for single-crystal work, in extension of the temperature ranges covered by the single-crystal thermostats previously described (Ubbelohde & Woodward 1943, 1946).

(2) To measure the thermal expansion and also the hybrid crystal formation accompanying the thermal changes at the Curie points, by X-ray methods.

(3) To obtain information about the nature of subcrystalline units in hybrid crystals, and about their merging into a single crystal above the transition point.

EXPERIMENTAL METHODS

(a) *Construction of a low-temperature micro-thermostat for single-crystal work.* As was shown in work with single crystals at high temperatures (Ubbelohde & Woodward 1943), it is comparatively easy to maintain a metal chamber practically as an enclosure at uniform temperature, provided that the chamber is not much larger than the crystal it surrounds and has walls of high thermal conductivity. The importance of keeping the chamber small is that this minimizes heat exchange between it and the X-ray camera, and also facilitates rapid changes of temperature when these are required.

Figure 1c illustrates the thermostat. A copper cup (1) is supported on a cork table by means of small wooden feet (not shown). This cup holds about 200 c.c. of liquid oxygen or other cooling medium, and is joined to a copper rod (4) which terminates in a short copper sleeve. The micro-thermostat itself consists of a copper rod drilled out at the bottom to form the constant-temperature chamber, and is a good push fit into this sleeve so that adjustments can be readily made to the position of the micro-thermostat in the sleeve. The object of this adjustment is to change the rate of heat transfer from the micro-thermostat to the cooling medium by changing the length of copper annulus along which this heat is conducted.

Details of the micro-thermostat are shown enlarged at the sides of figure 1 (a, b). The constant-temperature chamber has an internal diameter of 0.4 cm. and a volume of about 0.18 c.c., and is drilled out of rod 0.8 cm. diameter. A window (7 in figure 1c), 0.2 cm. high and 120° angular aperture, allows the entrance of the primary beam and exit of reflected beams with Bragg angles from 90° down to about 60° . The unreflected part of the primary beam passes on through a circular hole, 0.2 cm. in diameter, opposite the main window of the chamber.

In the present design of micro-thermostat these windows were covered with cellophane, since the smaller radiation exchange compared with the high-temperature

thermostat (Ubbelohde & Woodward 1943) made beryllium windows unnecessary; at very low temperatures beryllium windows would probably be desirable to improve black-body conditions within the chamber.

The lower portion of the chamber is wound with 'bright-ray' wire (about 5Ω total resistance), insulated from the copper chamber by means of very thin mica and lagged with silicate/asbestos cement.

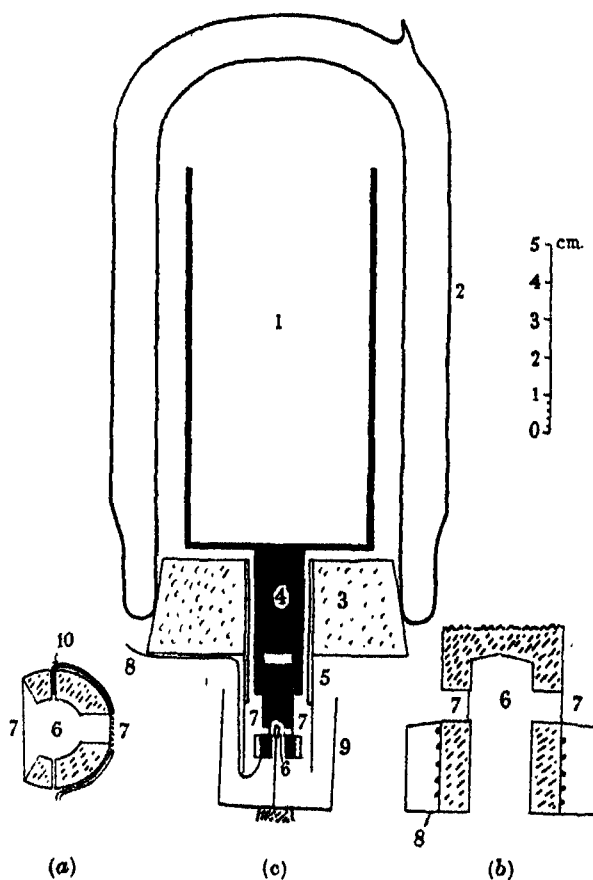


FIGURE 1. Low-temperature micro-thermostat. 1, copper container for cooling medium; 2, Dewar vessel; 3, cork table; 4, copper rod; 5, cellophane sleeve with glass stiffening; 6, chamber showing crystal in position; 7, cellophane windows; 8, leads for heating coil and thermocouple; 9, cellophane 'well' (mounted on crystal goniostat); 10, main thermocouple giving temperature of chamber walls. Figures 1 *a* and *b* give horizontal and vertical sections of the micro-thermostat, magnified four times.

A steady temperature is reached in the chamber when the flow of heat from the micro-thermostat to the copper sleeve, and the cooling medium, is balanced by the heat dissipated in the nichrome wire when a steady current is passed. Radiation exchange makes only a small contribution to the heat balance. Insulation of the

chamber from the outer air is obtained by means of a flow of cold dry gas down the outer walls. To supply this, a Dewar vessel (2) is inverted and fitted on to the cork table. As the cooling medium evaporates, the cold gas produced, which is roughly at the temperature of the cup and very dry, flows down a thin cellophane sleeve (5) which surrounds the copper rod and chamber. This sleeve is stiffened in its upper part, where this does not interfere with the X-rays, by means of a glass tube fitted into the cork. After issuing from the cellophane sleeve, the cold dry gas falls into a 'well' of cellophane stiffened with paper (9), and mounted round the crystal. From this well the cold gas overflows into the main camera.

By virtue of these counter-currents the most economical use is made of the cooling medium, and no frosting of the cellophane windows occurs, since the relative humidity of the gases in the neighbourhood is kept very low. A loose cloth cover checks entry of laboratory air into the main X-ray camera, which does not get cooled by more than about 5°C below room temperature, owing to the small quantity of heat exchanged by the micro-thermostat with its surroundings. For work near the boiling-point of the cooling medium, the micro-thermostat is pushed right home into the copper sleeve, and for work around 120°K it is pulled out by a few millimetres to increase the resistance to heat flow.

(b) *Assessment of temperature control in the micro-thermostat.* Temperature measurements of the walls of the chamber were made continuously, using a copper-constantan thermocouple calibrated at various fixed points. To eliminate leads corrections, 36-gauge wire was used, and the leads were wrapped half a turn round the walls of the chamber before taking them away, as shown in figure 1a (10).

A small auxiliary couple was used to investigate any temperature difference between the copper walls and the air space in the chamber. Corrections due to conduction of heat along the leads were avoided in a similar way as for the main couple, by pushing one or two loops of the lead wires well up into the air chamber, and then bringing the junction down to the position later occupied by the crystal. It was found that over the temperature range $95\text{--}130^{\circ}\text{K}$, the *maximum* gradient horizontally or vertically within the air chamber did not exceed 1°K/mm . When the heating current was altered so as to change the temperature, a new steady state was reached within 1 min. for changes of the order of 10 or 15°K . With no heating current, and the thermostat pushed home into the copper sleeve, the temperature was practically that of the cooling medium, and rose about 25°K for a dissipation of 15 W in the heating coil.

Temperatures below 85°K were not used for the experiments described below, but might readily be attained, e.g. by evaporating the cooling medium under reduced pressure.

In addition to using the thermostat when measurements were wanted over a range of temperatures, numerous observations were carried out at the boiling-points of oxygen and nitrogen.

(c) *Crystals* were prepared from aqueous solutions of A.R. salts saturated at about 60°C and allowed to cool slowly down to room temperature over a period of

about 12 hr. It seemed likely that changes in this procedure would affect the hysteresis (see below) in the thermal transformations, so that conditions were kept as nearly constant as possible for the present investigation.

For studying the (*hk*0) reflexions, clear needles were used approximately 0.3 mm. thick. For (*Ok*l) reflexions, pieces were cleaved from the ends of rather larger crystals. Some observations were also made on the (*hkh*) zone. The crystals were mounted at the ends of thin glass fibres by means of nitrocellulose dissolved in acetone + amyl acetate, and were enclosed in thin cellophane tubes when subjected to drips of liquid oxygen or nitrogen. In view of the transformations described below, it was noteworthy that such crystals retained their brightness of faces and sharpness of Webski reflexions, even after many cycles of cooling and heating.

(d) *X-ray methods.* A multiple-exposure camera (Ubbelohde 1939*a*) was used for most of the Bragg and Laue photographs, and a few exposures of the hybrid crystals were also taken on a moving-film spectrometer (Robertson 1934). Unfiltered Cu *K* α radiation was used for most of the Bragg photographs, and unfiltered Mo *K* α radiation to test a special point (see below). Laue reflexions were obtained down to wave-lengths of about $\lambda = 0.4$ Å, using a copper anticathode; the collimator holder used allowed both equatorial reflexions and those for which the reflected ray was inclined at angles up to about 30° to be recorded on the film (cf. Ubbelohde 1939*a* and figure 4, plate 12).

EXPERIMENTAL RESULTS

As stated above, it was observed for KH_2PO_4 that at low temperatures certain reflexions were split; a similar effect of about the same magnitude was confirmed in the case of KH_2AsO_4 , but detailed investigations on this crystal were not pursued for the purpose of the present investigation. In this connexion, reference may also be made to effects recorded by Swiss workers, which must be due to hybrid crystal formation (Bantle 1942; Quervain & Zwicker 1943; Quervain 1944).

The change in crystal structure giving rise to this split was found to be reversible, in the sense that on raising the temperature again single reflexions were observed once more, either immediately or after a variable lapse of time. In view of this thermal transformation in the crystals, it is convenient to consider the thermal contraction down to just above the transition point (about 123° K) separately from the changes occurring during the thermal transition and at still lower temperatures.

(a) *Thermal changes above the Curie point for KH_2PO_4*

Lattice spacings were derived using a number of reflexions from (*hk*0) planes, calibrated as previously described (Ubbelohde 1939*a*). Values were plotted against $\cos^2 \theta$ to give a nearly linear extrapolation to $\theta = 90^\circ$. On the basis of $\Delta g = 4.0775$ Å at 20° C, calculated axial lengths are

$$a = b = 7.437 \pm 0.002 \text{ Å.}$$

For the c axis, only two planes (048 with $\theta = 78^\circ$ and 093 with $\theta = 81^\circ$) could conveniently be used to give accurate values

$$a = b = 7.434 \text{ \AA}, \quad c = 6.945 \text{ \AA}.$$

All three axial lengths for tetragonal KH_2PO_4 agree within experimental error with the less precise data previously published (West 1930).

TABLE 1. THERMAL EXPANSIONS OF TETRAGONAL KH_2PO_4 ABOVE THE CURIE POINT

temp. range °K	crystal plane	contraction $\times 10^3$	remarks
		Å	
293/123	028	5.73	mean of 14 independent observations
	048	5.46	mean of 14 independent observations
	084	4.05	mean of 11 independent observations
	480	3.75	mean of 14 independent observations

Mean thermal expansions calculated from the above data are:

$$\begin{aligned} \Delta a = \Delta b &= (2.73 \pm 0.05) \times 10^{-2} \text{ \AA}, & \alpha_{11} = \alpha_{22} &= 2.16 \times 10^{-5} \text{ per } ^\circ\text{K}, \\ \Delta c &= (4.03 \pm 0.05) \times 10^{-2} \text{ \AA}, & \alpha_{33} &= 3.43 \times 10^{-5} \text{ per } ^\circ\text{K}. \end{aligned}$$

Effect of temperature on intensities of reflexions. As described for Rochelle salt (Ubbelohde & Woodward 1946), estimates of the ratio of intensities at two different temperatures can be readily made, by means of the multiple exposures on the same film. Such intensity ratios can only be interpreted in terms of atomic vibrations within the crystal, provided the degree of perfection of the crystal and consequently the extinction is not seriously altered as a result of successive cycles of change from single to hybrid crystal and back again.

Visual estimates of intensity ratios for the $(0kl)$ reflexions showed that mean values referred to room temperature, and a temperature either just above or just below the Curie point, agreed within 8% for the 028 and 048 planes, i.e. that any effect of change of extinction on passing through the Curie point was not more than a subsidiary factor for reflexions in this zone. If the intensity ratio was ascribed entirely to the change in the temperature, then, assuming the standard Debye-Waller expression (cf. Ubbelohde & Woodward 1946), the characteristic temperature for vibrations parallel to the c axis was about $244 \pm 6^\circ \text{K}$, from the results of two independent observers. This would give a root-mean-square amplitude of vibration of about $0.097 \pm 0.002 \text{ \AA}$ at 123°K and $0.143 \pm 0.003 \text{ \AA}$ at 293°K , parallel to the c axis.

On the other hand, reflexions in the $(h k 0)$ zone gave widely varying intensity ratios in different experiments, values differing by as much as 300% in different comparisons between room temperature and temperatures just above or just below the transition point. This indicates that the degree of perfection or mechanical strain in KH_2PO_4 can vary notably in different directions in the crystal; further work is in hand on this point, in view of its interesting bearing on the mechanical behaviour of crystals.

(b) *Thermal changes in the neighbourhood of the Curie point for KH_2PO_4 lattice spacings*

When crystals of tetragonal KH_2PO_4 were suddenly chilled from 293° K, by dripping liquid oxygen (90° K) or liquid nitrogen (77° K) on to them, certain reflexions split. As shown by detailed studies summarized below, this was due to a transformation of the single crystal into a hybrid made up of subcrystalline units whose lattice dimensions differ only to a small extent from those of the tetragonal crystal lattice. Under these conditions of sudden cooling, the change in lattice dimensions appeared to be quite reproducible and was practically the same at 90° K as at 77° K. From fifty-seven independent observations on eight different planes in the (*h**k*0) zone it was found that

$$\Delta a = \Delta b = (2.45 \pm 0.05) \times 10^{-3} \text{ \AA}, \quad \frac{\Delta a}{a} = \frac{\Delta b}{b} = (3.29 \pm 0.06) \times 10^{-3},$$

and from thirty-one independent observations on six planes in the (0*kl*) zone, and the above values of Δa and Δb ,

$$\Delta c = (4.13 \pm 0.02) \times 10^{-3} \text{ \AA}, \quad \Delta c/c = (5.95 \pm 0.02) \times 10^{-3}.$$

Although these changes are not, strictly speaking, thermal expansions of any one crystal lattice, the structure of the sub-units in the hybrid resembles the original tetragonal structure so closely that the above contractions bear a fairly direct relation to either structure.

Attempts were also made to measure the true thermal expansion of the sub-units in the hybrids by using the micro-thermostat over the range 128 to 95° K. However, it proved difficult to obtain consistent numerical values in this region. As in the case of Rochelle salt (Ubbelohde & Woodward 1946), hysteresis was observed just below the Curie point, and the behaviour of the hybrid crystal was governed by factors in its previous history, such as the time during which it was maintained in the immediate neighbourhood of the transition point before cooling further (see below). Observations of interest included:

(i) The angle between the axes of the subcrystalline hybrids, as measured by the separation of components in a split reflexion, in the suddenly chilled crystal, was $90^\circ 26.85' \pm 0.17'$ at 90° K (see below). It decreased as the crystal was warmed to within 3 or 4° of the transition point, as was shown by the split reflexions coming somewhat closer together. At slightly higher temperatures very rapid coalescence occurred to a single fuzzy reflexion. The maximum decrease in angle between *a* and *b* axis of the sub-units which could be clearly measured before this coalescence occurred was from $90^\circ 27'$ to about $90^\circ 13'$. Although the split reflexions were quite sharp at 77 and 90° K, they frequently exhibited diffuse wings even before they coalesced in the neighbourhood of the transition point, showing that the sub-units present might be in a state of very considerable strain.

(ii) Changes of length observed on continuous cooling of a crystal are recorded in table 2.

TABLE 2. CHANGES IN AXIAL LENGTHS OF KH_2PO_4 ON COOLING THROUGH THE TRANSITION POINT

temperature $^{\circ}\text{K}$	Δa	Δc (in Å)
293	($a = 7.437$)	($c = 6.945$)
173	-0.0212	-0.0317
123	-0.0273	-0.0405
118	-0.0229	-0.0405
90	-0.0245	-0.0413

With reference to the above results, it may be noted that the a (and b) axes *increase* in length on cooling below the transition point, and then decrease slightly on cooling still further. Corresponding changes have recently been observed by optical means on the macroscopic scale (von Arx & Bantle 1944). Rather surprisingly, the optical values (shown as a full curve in figure 2) for thermal contraction when the crystal was cooled very slowly, which make no allowance for hybrid formation, agree

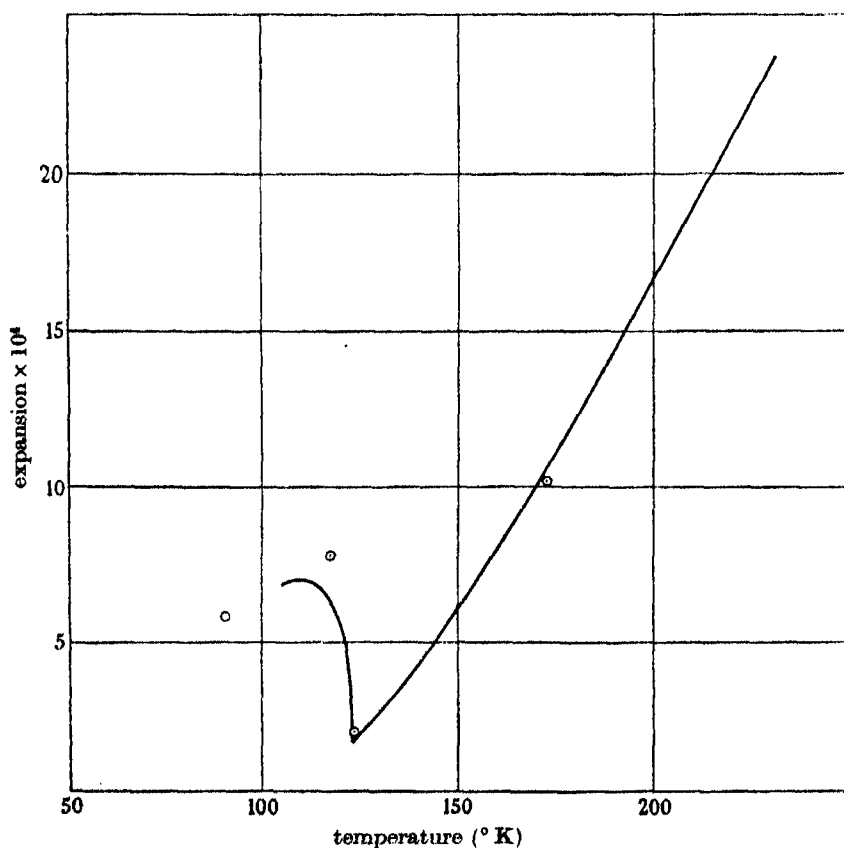


FIGURE 2. Comparison of X-ray and optical measurements of contraction of a axis of KH_2PO_4 on cooling. \odot X-ray data, — optical data.

with the X-ray results so far obtained for the a or b axis of the crystal lattice cooled fairly rapidly (shown as points in figure 2; a common zero has been assumed at 293° K in comparing the data). This conclusion may need revision when more data are available concerning the effect of rate of cooling on the supercooled state (see below).

No measurable change in length of the c axis could be detected accompanying the thermal transition from 123 to 118° K. Below this temperature the contraction was less rapid than would be extrapolated from data above the transition point. Some care was taken to test whether the c axis of subcrystalline units was at all deflected in a random way from the original tetragonal c axis. A faint ghost reflexion observed near the strong 048 might indicate a tilt of not more than 4', or an expansion of the c axis by not more than 0.15 %, in certain parts of the crystal, but this indication should not be regarded as substantiated till further evidence is available.

(c) *Nature of the hybrid crystals*

(i) *Lattice dimensions of the sub-units.* Measurements were made on crystals chilled to 90° K, using rotation and oscillation photographs for ($hk0$) and ($0kl$) zones, and for some (hkh) reflexions. Splitting of the reflexions from the tetragonal crystal was observed whenever $h \neq 0$ and $k \neq 0$, and the separation of the split components was largest when $h = k$. No difference was observed between reflexions of the types (hkl) and ($\bar{h}kl$). Moving-film rotation photographs confirmed the above distribution, which was established by oscillation photographs.

The separation Δ of the split components was measured on eight independent reflexions, with Bragg angles ranging from 48 to 79°, and values of $(\Delta/d) [(h^2 + k^2)/2hk]$ calculated from these measurements showed no systematic variation with lattice spacing d . Furthermore, no splitting was observed in reflexions for which $h = 0$ or $k = 0$. It was deduced that the a and b axes of the original tetragonal cell remain equal in length in the sub-units, but that the angle between these axes differs slightly from 90° in the sub-units.

From forty-nine observations on seven different planes, this angle was calculated to be $90^\circ 26.85' \pm 0.17'$ at 90° K. At 77° K rather fewer observations gave $90^\circ 27.2'$.

As explained above, there was no clear evidence of split ($0kl$) reflexions; the c axis of the sub-units remains normal to the plane containing the a and b axes, at least within $\pm 2'$.

To test whether sub-units packed with a different degree of strain in the interior of the crystal from their packing in the surface, some observations were made with Mo $K\alpha$ radiation. Calculations showed that this is nine times less absorbed than Cu $K\alpha$ radiation, and that a crystal 0.3 mm. thick would absorb less than half the incident Mo $K\alpha$ radiation. Suitable back reflexions were not available with this wave-length to permit very accurate measurements, but by 'bracketing' the split of the 1060 reflexion between those of 840 and 8100, and comparing it with the separation of the $\alpha_1\alpha_2$ reflexions on the multiple exposure film, the angle calculated

was $90^{\circ} 27.3' \pm 1.4'$ at 90° K, which verified that sub-units in the interior of the crystal were substantially in the same state as those in the surface.

(ii) *Mutual arrangement and size of sub-units in a split crystal.* In Bragg photographs, both components of a split reflexion for h or $k \neq 0$ were observed in all cases, with approximately equal intensities, when the crystals were oscillated through 15° .

More detailed information about the mutual arrangement and size of the crystallites was obtained from Laue exposures. In some of these the needle-shaped crystal, mounted with the c axis vertical, was only about 0.25 mm. broad and the cross-section of the X-ray beam was 0.7 mm. broad and 1 mm. high. Very good definition was obtained, even with components only 0.4 mm. apart on the film.

In these Laue photographs, reflexions from crystals at 90° K were in general found to be split into four components (figure 4, plate 12) corresponding with two pairs of distinct orientations of sub-units, all having the c axis parallel to the original c axis of the single crystal. The separation between the pairs of components varied systematically with h/k . As h/k tended to unity, the outer pair of components approached the inner pair, and when $h = k$ they coalesced, leaving only two distinct reflexions. As h/k tended to zero or infinity, the inner pair of components approached, and coalesced when $h/k = 0$ or ∞ , leaving only three distinct reflexions.

Quantitatively, the separations in a pair of components agreed with the expressions $2\gamma[1 + h^2/k^2]$ for one pair, and $2\gamma[1 + k^2/h^2]$ for the other, where

$$\gamma = 26.85' = 1.095 \text{ mm.}$$

on the film. From visual observations, the intensities of the two outer reflexions and the two inner reflexions were found to be respectively equal to one another, but the intensity of one pair might differ from the other by a factor of as much as ten. Further, one pair might be displaced relative to the centre of the other, in a horizontal or vertical direction, or both. (In one film a displacement of 0.15 mm. was noted.)

From the components observed in split reflexions, it was concluded that when the tetragonal crystal changes to a hybrid, the subcrystalline units of lower symmetry assume only a limited number of directions with respect to the original framework, and spring from it so as to retain *either* the a or the b axis in common with the original tetragonal framework, and the c axis parallel to that of the original single crystal. The four possible arrangements on this scheme are illustrated in figure 3. Pairs of sub-units numbered 1 and 2, or 3 and 4, have the polarity of their c axes pointing in opposite directions.

In such a crystal, the plane ($hk0$) has in general a different orientation in each of the four crystallites, and so gives rise to a Laue reflexion split into four components; as the orientations of the (100) coincide in 3 and 4, and those of the (010) in 1 and 2, there are for these planes three orientations only and so three components; while for the (110) the orientations coincide for 1 and 3, and also for 2 and 4, giving two components of the split Laue reflexion.

From the observed intensities, each member of a pair occurs with about the same probability within the area of crystal illuminated by the X-ray beam, i.e. 2.5×10^{-3} sq.cm. Such statistical smoothing would hardly occur unless there were say at least one hundred sub-units within the cross-section of the beam, so that an upper limit for the size of the units which give rise to a pair of reflexions, if the arrangement is random, would be about 5×10^{-3} cm. edge. Each kind of pair appears to tend to aggregate with neighbours of the same kind, to explain the differences of intensities between pairs.

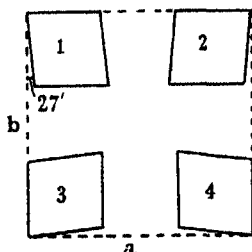


FIGURE 3. Section of crystal hybrid normal to c axis, showing four possible modes of genesis of sub-units.

(iii) *Rate of transformation from a single crystal to a hybrid and back.* A complete study of the transformations within the solid state, which can now be made by the means described above, would be a large undertaking. The 480 reflexion was the strongest with Bragg angle sufficiently large (68°) to give a clear splitting of the components from the hybrid, and the following briefly illustrates the kind of observations made up to the present in a series of oscillation photographs made between 128° and 100° K. Starting with the crystal hybrid formed by chilling KH_2PO_4 down to about 100° K, the components of the split reflexion were observed to come nearer to one another as the crystal was warmed to within 3 or 4° below the transition point. At slightly higher temperatures they coalesced very rapidly, leaving a fuzzy single reflexion. In terms of the angle between the a and b axes of the sub-units, this decreased gradually from about $90^\circ 27'$ to $90^\circ 13'$, and then rapidly to very approximately 90° apart from local strains.

Starting with the crystal just above the transition temperature, if the crystal was maintained near the transition temperature for some time, then on cooling to 100° K the separation was appreciably below the normal value observed on rapid chilling, and in a few cases the reflexion appeared to remain single, occupying approximately the mean of the positions in which the split components might be expected. In this region of temperatures the detailed structure of the hybrid appears to be subject to hysteresis.

On warming the hybrid crystal back to room temperature, sharp single reflexions were obtained after a short lapse of time. The detailed measurement on the rate of reversion from hybrid to single crystal at intermediate temperatures above the transition point is still under investigation.

DISCUSSION

At the present stage of this investigation it will be sufficient to draw attention to the way in which the results link up with other investigations on the solid state.

(i) *Terminology of hybrid crystals.* The only reason for the new name 'hybrid' crystals is that none of the standard nomenclatures completely fits the case.

According to standard space-group theory, single crystals of KH_2PO_4 (and KH_2AsO_4) which are stable at room temperature are tetragonal sphenoidal, with space group

$$D_{2d}^{13}-\bar{4}2d(a).$$

Strictly speaking, this space group applies only to ideal crystals. More or less random mosaicity introduces a slight indeterminacy or blurring in the exactness with which the symmetry operations can be applied to the crystal as a whole, and such mosaicity may be favoured for thermodynamic reasons (Oldham & Ubbelohde 1940).

It seems likely that the individual sub-units of a hybrid crystal can also be described by standard space-group theory. Compared with the original tetragonal lattice of KH_2PO_4 , the dyad and dyad screw axes parallel to the a and b axes disappear, and the fourfold alternating axis parallel to the c axis becomes a dyad axis. Symmetry elements retained are dyad and dyad screw axes parallel to the c axis, and glide planes parallel to the diagonals of the cell. Choosing these diagonals as the new a and b axes, and doubling the size of the unit cell, the symmetry elements are those of the orthorhombic hemihedral space group $C_{2h}^{19}-Fdd$.

Such geometrical space-group descriptions do not, however, completely describe the physical reality of the hybrid crystal. In the first place, the individual sub-units have a definite geometrical and statistical relationship to each other and to the original framework. The close association between pairs of units of opposite polarity has certain analogies with twinning, though the thermal origin of the change introduces a new factor in the formation of this state. Secondly, in the immediate neighbourhood of the transition point, adjoining regions of the crystal must be in a state of strain, as shown by the diffuseness of the Bragg reflexions, and by the different orientations of sub-units shown by certain Laue photographs. In this condition the sub-units deviate appreciably from the geometrical relationships of the standard space group.

Extending these considerations to other crystals, it may prove necessary to replace geometrical ideas of space-group operations with exact identities by physical ideas in which any operation gives identity within a certain indeterminacy. In fact, the problem to be solved experimentally is what degree of indeterminacy actually applies to any given space-group operation, when referred to a real crystal. Sub-units in a hybrid with differences smaller than those described above might escape experimental detection, and yet be significant in determining the equilibrium of the lattice.

(ii) *Phase rule interpretation of continuous thermal transitions.* According to the classical derivation of the phase rule, a transition point in a one component system

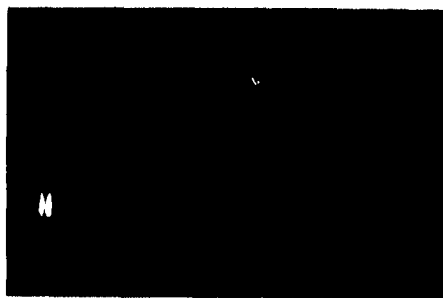
involves the co-existence of three independent phases, with the same thermodynamic potential of the component in three independent equations of state. For equilibria involving two solids and one vapour phase, the requirement that the phases shall be 'independent' can be given a definite meaning in terms of crystal structure. For example, if we start with a single crystal of β -resorcinol, and cool it below the transition point with α -resorcinol, the crystal may appear to preserve its external form, but X-ray photographs show that actually it breaks up into a mass of smaller crystals of the new variety. In so far as these are independent of the original crystal lattice, they give powder rings corresponding with the new crystal lattice, i.e. they are distributed at random with respect to the original lattice (Robertson & Ubbelohde 1938). Accompanying this rearrangement of the molecules to form a new crystal structure independent of the original arrangement, there are finite changes of volume and heat content at the transition *point* between the two crystal forms. Such transitions have sometimes been termed 'of the first kind'. We propose to term them 'discontinuous'.

It has been pointed out earlier (Ubbelohde 1937*a, b*) that if an additional parameter such as an internal strain, or state of order, is required to describe the thermodynamic potential of say one of the solid phases, an additional degree of freedom is introduced. Equilibrium may be established with reference to micro-fluctuations in the system, but there is hysteresis owing to the low probability of 'co-operative' or macro-fluctuations.

Experimental examples of thermodynamic transitions termed 'of the second kind', which we propose to term 'continuous', are fairly numerous (Ubbelohde 1940, 1941). In these, finite changes of volume and heat content at a transition point are replaced by 'anomalous maxima' on the expansion and specific heat curves, suggesting that the change is spread out over a range of temperatures. Although some of these phenomena have been analysed by statistical mechanics, it has not hitherto been clear how such continuous transitions affect crystal structure.

The formation of hybrid crystals in the transformations now described for Rochelle salt and KH_2PO_4 make it clear that a break-up of the single crystal actually does occur over the transition range of temperatures. But apparently the change in structure involved is sufficiently small to be accommodated within the original lattice by the formation of a hybrid in which the units are oriented in definite relationships to the original crystal. When this can happen, the mutual strains imposed on the sub-units and the condition for their packing mean that we no longer have two independent equations of state for the two crystal lattices, and the thermodynamic change becomes continuous and subject to hysteresis.

Other examples of continuous thermodynamic transitions can probably be explained in the same way. Examples in which the experimental evidence is fairly clear include (*a*) the pre-melting of long-chain hydrocarbons, where there is close similarity between the solid and liquid phases, and the transition range covers several degrees (Ubbelohde 1938); (*b*) a phenomenon similar to the formation of hybrid crystals which has recently been demonstrated in the transition in am-



(a)

(b)

FIGURE 4. Laue photograph of a KH_2PO_4 crystal under different conditions. (a) Comparison of Laue reflexions with same orientation of crystal below (upper photograph) and above (lower photograph) transition point. (b) Laue reflexions below transition point with different orientations of crystal, showing components with different intensities.

monium chloride at low temperatures (Dinichert 1942, 1944; Extermann & Weigle 1942); (c) a similar phenomenon has now been observed in a thermal transition in strontium-barium titanate (Megaw 1946). It should be pointed out, that although the distinction between a single crystal and a hybrid is a fundamental one in the thermodynamic sense, it requires comparatively refined X-ray methods to establish the existence of hybrid crystals. It is conceivable that, for example, large organic crystals such as proteins are hybrid, and this would have an important bearing on the interpretation of the functions of these molecules in relation to their crystal structure.

Discussion of the thermal expansion in relation to the structure of KH_2PO_4 is deferred to a subsequent paper.

Thanks are due to the Managers of the Davy Faraday Laboratory of the Royal Institution for providing facilities for this work.

The Swiss papers subsequent to 1939 have only very recently come to our notice, and we would like to thank Professor Weigle of Geneva for informing us about this work, which links up with that described above, but which owing to war conditions was not available when our experiments were carried out.

REFERENCES

- von Arx & Bantle 1944 *Helv. Phys. Acta*, **17**, 299.
Bantle 1942 *Helv. Phys. Acta*, **15**, 373.
Busch 1938 *Helv. Phys. Acta*, **11**, 269.
Dinichert 1942 *Helv. Phys. Acta*, **15**, 462.
Dinichert 1944 *Helv. Phys. Acta*, **17**, 389.
Extermann & Weigle 1942 *Helv. Phys. Acta*, **15**, 455.
Megaw 1946 *Nature*, **157**, 20.
Oldham & Ubbelohde 1940 *Proc. Roy. Soc. A*, **176**, 50.
Quervain 1944 *Helv. Phys. Acta*, **17**, 509.
Quervain & Zwicker 1943 *Helv. Phys. Acta*, **16**, 216.
Robertson 1934 *Phil. Mag.* **18**, 729.
Robertson & Ubbelohde 1938 *Proc. Roy. Soc. A*, **167**, 138.
Ubbelohde 1937*a* *Trans. Faraday Soc.* **33**, 1203.
Ubbelohde 1937*b* *Proc. Roy. Soc. A*, **159**, 301.
Ubbelohde 1938 *Trans. Faraday Soc.* **34**, 292.
Ubbelohde 1939*a* *J. Sci. Instrum.* **16**, 155.
Ubbelohde 1939*b* *Proc. Roy. Soc. A*, **173**, 417.
Ubbelohde 1940 *Ann. Rep. Chem. Soc.* **36**, 157.
Ubbelohde 1941 *Ann. Rep. Chem. Soc.* **37**, 174.
Ubbelohde & Woodward 1943 *Proc. Roy. Soc. A*, **181**, 415.
Ubbelohde & Woodward 1945*a* *Nature*, **155**, 170.
Ubbelohde & Woodward 1945*b* *Nature*, **156**, 20.
Ubbelohde & Woodward 1946 *Proc. Roy. Soc. A*, **185**, 448.
West 1930 *Z. Kristallogr.* **74**, 306.

The mechanism of corrosion fatigue of mild steel

BY U. R. EVANS AND M. TCHORABDJI SIMNAD

(Communicated by H. J. Gough, F.R.S.—Received 23 February 1946)

The object of this research was to establish the mechanism of corrosion fatigue by the aid of chemical and electrochemical measurements, with special reference to the possibility of preventing corrosion fatigue by means of cathodic currents.

Two-stage tests on steel specimens subjected to alternating stress, with a chloride solution applied during the first stage only, have indicated that as the first stage (corrosion period) is increased, the total life, at first immeasurably long, becomes extremely short and then increases again. The unexpected increase in total life produced by an extension of the corrosion period may be explained by the fact that isolated cracks produce more stress intensification than a number of neighbouring cracks. Another series of experiments has shown that the rate of passage of iron into the combined state greatly increases with the applied stress.

The application of a cathodic current diminishes the rate of production of iron compounds and the number of cracks; weak cathodic currents actually shorten the life, but still stronger ones increase it again—which can be explained in two ways: (1) for a given depth few cracks cause more weakening than many, (2) for a given amount of corrosion, an increased number of cracks means smaller depth of cracking and hence less damage. If the current reaches a certain value, corrosion becomes undetectable, and the life becomes extremely long in neutral potassium chloride (but not in acid). The value of the current needed for this complete protection increases with the stress range, but the value of the potential corresponding to the protective current moves steadily lower, i.e. in the direction of zinc, with applied stress.

Applying the graphical methods of representing corrosion phenomena to these results, we are led to the view that at least three different factors operate in causing alternating stress to enhance the rate of corrosion and the rate of mechanical damage. These are (A) diminution of cathodic polarization, (B) diminution of anodic polarization, and (C) diminution of the resistance of the path joining anodes and cathodes. It is possible that there may also be (D) a bodily shift of the anodic polarization curve in the base metal direction.

Studies of the shift of potential with time in presence of different types of stress indicate that stressing within the elastic range only affects the potential by altering the state of repair of the film covering the surface. It is likely that stresses within the plastic range depress the potential of the metal itself—irrespective of any damage to a film—but further work will be needed definitely to establish this point.

INTRODUCTION

Earlier researches. Corrosion fatigue, discovered by Haigh (1917), was studied in great detail by McAdam (1926–31). The experimental data were brought together by Gough (1932), who associated the enhanced damage produced by superimposing cyclic stresses on corrosive action with the continued breaking of films; Bengough (1932) attributed it to an increased supply of oxygen.

Gough & Sopwith (1932, 1934), during a microscopic and X-ray examination of aluminium, observed (1) many small pits, (2) local attack with larger pits, and (3) preferential attack on slip-bands, leading to cracking. Gough (1932, 1935) obtained an enhanced life by working *in vacuo*. The peripheral extension of the cracks was studied by Bacon (1935), whilst Gould's (1934) measurements provided further

evidence that failure is due to sharp-ended cracks rather than rounded pits—a distinction well brought out in the photographs of McAdam & Clyne (1940).

The inhibitive action of chromates was demonstrated by Speller, McCorkle & Mumma (1928), and studied quantitatively by Gould & Evans (1939). The latter also measured the increase in life obtained by cathodic action, such as is produced by contact with zinc—an effect demonstrated in earlier work by Behrens (1933), Haigh (1929), Gerrard & Sutton (1935), Krystof (1935), and later studied by Stuart & Evans (1943), and by Huddle & Evans (1944), who used a paint richly pigmented with metallic zinc powder (Mayne & Evans 1944).

Plan of new research. The investigations mentioned above provided much information on the mechanical and crystallographic aspects of corrosion fatigue, but little regarding the chemistry. Doubt remained as to whether alternating stress increased the total rate of destruction or merely redistributed attack so as to produce greater mechanical weakening. In the present research, modern photo-electric methods of microanalysis have been used to obtain an answer to this question.

The influence of alternating stress upon corrosion has been attributed by different investigators to three main groups of factors:

(I) The distortion or obliteration of the crystalline structure of the metal, which thus becomes less stable and more reactive.

(II) The rupture of protective films, or removal of corrosion products which would otherwise slow down the attack.

(III) The improvement in the supply of oxygen, which is needed for the corrosion.

A further object of the present research was to decide between these three explanations.

The matter can be restated in an electrochemical form. It has in recent years been established that corrosion by most salt solutions follows an electrochemical course, since electric currents flowing between discrete anodic and cathodic areas have been detected on naturally corroding specimens, and have been found strong enough to account for the corrosion actually observed. Translated into electrochemical nomenclature, the factors suggested above become:*

(1) a bodily shift of the anodic potential in the 'base-metal' direction;

(2) (a) diminished anodic polarization, or (b) diminished resistance of the electrolyte path between anodes and cathodes;

(3) diminished cathodic polarization.

Since cathodic stimulation tends to raise the compromise potential of a corroding system, and anodic stimulation to depress it, the measurement of the electrode potential of metal subjected to corrosion fatigue should serve to distinguish between factors (1) and (2) on the one hand and (3) on the other.

In applying the electrochemical methods just indicated to elucidate the mechanism of corrosion fatigue, some information has also been obtained in regard to the

* Broadly speaking (1), (2), (3), are restatements of (I), (II), and (III) respectively, but Dr J. N. Agar has kindly pointed out that (I) could give rise to (2a).

question as to whether stressing within the elastic or plastic range alters the electrode potential of a metal. Such questions possess a fundamental importance—quite apart from corrosion fatigue.

EXPERIMENTAL METHODS

Materials. The specimens were wires 15 in. long and 0.1 in. diameter, produced from a single billet of mild steel by Messrs Brunton's, Ltd.; drawing was carried out in two passes from patented steel rod 0.144 in. diameter. The ultimate tensile strength was found to be 48 tons/sq.in., and the fatigue limit of wires coated with vaseline containing zinc chromate ± 21 tons/sq.in. Chemical analysis, kindly arranged by the late Dr W. H. Hatfield, F.R.S., appears in table 1; microscopical examination showed freedom from inclusions.

TABLE 1. CHEMICAL COMPOSITION OF SPECIMENS

element	content %
carbon	0.19
silicon	0.07
sulphur	0.03
manganese	0.78
phosphorus	0.02

The air-fatigue properties of such cold-drawn wire have been studied in detail by Gill & Goodacre (1936), Shelton & Swanger (1935), and by Godfrey (1941).

Apparatus. The experiments were performed in a room thermostated at $25 \pm 2.5^\circ\text{C}$, since Gould (1936) had shown that the effect of larger temperature fluctuations is not negligible in corrosion fatigue. A Haigh-Robertson wire-fatigue machine was adopted for its high frequency of stressing and suitable size. A wire bent into bow form is made to rotate about its own axis of flexure, being so gripped and loaded that fracture occurs very near to mid-span, where the bending stress is maximal, namely $\pi E\theta d/2L$ for a specimen of length L , diameter d , angle of flexure θ , and elastic modulus E .

The machine was carefully earthed to prevent interference by stray currents. A counter was fitted to the motor for recording the number of cycles, and a check was kept on the speed of the motor by means of a stroboscope disk and an electric clock. Fluctuations on the mains voltage were corrected by using a 'Raytheon' voltage regulator in the main circuit, whilst a 'Variac' auto-transformer was used to control the speed of the motor.

The feeding arrangement. None of the methods used hitherto seemed suited to this research, since they either give no well-defined margin and flow of liquid to the wetted area, or interfere with the products of corrosion. The arrangement developed for feeding the corrosive liquid and applying a cathodic current is shown in figure 1. The solution is allowed to siphon from a reservoir into the glass tube A and down the vertical glass rod B at a rate of 100 c.c./hr. The rod B , 5 mm. diameter, is waxed

completely except for an unwaxed channel along the side nearest to the wire. From a distance of about 1 cm. above the specimen, liquid descends along this unwaxed channel and a continuous ring of liquid is formed around the wire, held by capillary action. The liquid then descends into the horizontal tube *C* to which is fused a platinum foil anode. When a current is to be applied to the specimen, a high external voltage is used in conjunction with a high external resistance, so that small changes of resistance in the liquid portion of the circuit do not appreciably affect the current; the distance between the specimen and the platinum anode is kept constant. The liquid, flowing out of *C* through an opening at one end, is collected in a beaker for estimation of iron.

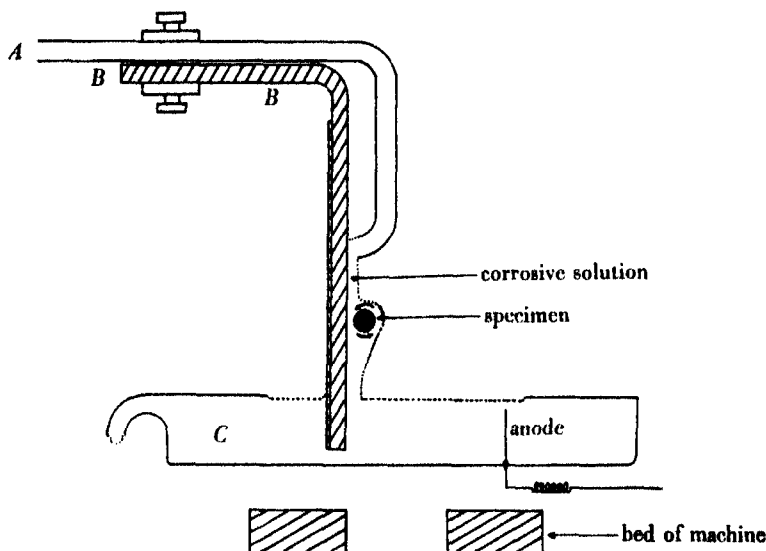


FIGURE 1. The feeding arrangement.

General procedure. The specimen was tested for straightness, ground with 000 emery over a length of 2.5 cm. at mid-span, and degreased with acetone. Then polystyrene lacquer was applied to the whole of the ground length, except for a length of 0.5 cm. at its centre, so as to confine the corrosion to this area. A period of 60 min. always elapsed between the grinding and the beginning of the test.

The specimen was placed in the machine and the speed of the motor adjusted to 6000 r.p.m., a speed chosen on account of the conditions under which the rotating specimens became dynamically unstable (large vibrations occur near the natural frequency of lateral vibration, and a compromise was necessary between the highest speed possible for steady flow of liquid around the specimen and the critical value). The knurled head of the tailstock was turned until the headstock had moved through the angle corresponding to the required stress range, which was always well below

the air-fatigue limit. The frequency of stressing was adjusted, and the rod *B* and tube *A* were moved towards the specimen until the liquid flowing down the rod made contact with the specimen at mid-span. The counter was then read and the time recorded. On termination of the test, the time and the number of cycles endured were noted.

Except where otherwise stated, the corrosive liquid was $M/10$ potassium chloride shaken with air before use. The pH varied from 6.06 to 7.0; according to Fink, Turner & Paul (1943) changes of pH within still wider limits are of no consequence in corrosion fatigue.

The iron passing into the combined state (including any rust on the specimen, which was wiped off with filter paper) was estimated by a colorimetric method based on thioglycollic acid. The use of a Spekker photoelectric absorptiometer made possible the determination of quantities of iron as small as 0.005 mg.

Special procedures. In *two-stage experiments*, where corrosion was to take place only during the first stage, the flow of the corrosive was stopped after the chosen period and the rust on the specimen collected by wiping with filter paper. Next, while the specimen was kept rotating at a slow speed, the corroded area was well washed with distilled water and then was swabbed repeatedly with cotton-wool soaked in molar potassium chromate. The final rubbings were carried out with dry cotton-wool and filter paper. The experiment was then continued in air until fracture occurred.

Where it was desired to measure the *electrode potential* during a test, a saturated calomel electrode was put into communication with the specimen by means of a filter-paper strip, supported by the waxed rod *B* through a bridge of decinormal potassium chloride (figure 2). The electrical connexion consisted of a small copper brush which made contact with the chuck of the motor and led, in conjunction with a connexion from the standard electrode, to a Cambridge Valve Potentiometer. The

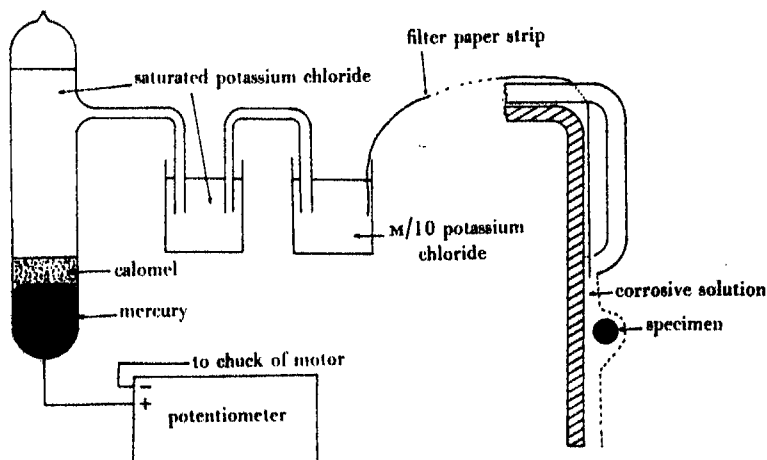


FIGURE 2. Arrangement for electrode-potential measurements.

potentials could be read to within 0.5 mV, and the first reading could be taken about 30 sec. after the liquid had made contact with the specimen.

Apparatus for static stresses. Where it was desired to measure the potential on non-rotating specimens subjected to static stresses, a special apparatus was designed to keep a wire specimen bent into the form of an arc, as in the Haigh-Robertson machine. The bending stress was, however, measured in terms of the deflection Y at mid-length, being $\pi^2 E Y d / 2 L^2$, where d is the diameter of the wire, L the length of the wire, and E Young's modulus.

The apparatus is shown in figure 3. The fixed headstock A contains a Haigh-Robertson bearing B consisting of four ball bearings. The thrust exerted by the wire is transmitted through a sleeve to a single 5 mm. ball that runs on a trio of balls of the same size. The latter are contained inside a race in A . The tailstock consists of a threaded cylinder C , held in the threaded holes of the fixed support S . At the end B' of C is a ball bearing, identical with that in A , for holding the other end of the specimen. In order to load the specimen the tailstock is rotated until the specimen is flexed to the desired extent by the longitudinal compression.

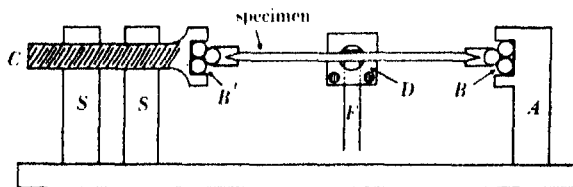


FIGURE 3. Apparatus for static stress experiments.

Specimens were ground 24 hr. before each experiment over a mid-span length 2.5 mm., and coated with polystyrene lacquer except for a central area 1 mm. wide by 5 mm. long, the longer side being parallel to the wire axis. A strip of filter paper 1 cm. wide, wetted with M/10 potassium chloride, made contact with this exposed surface, and was kept in position by means of the special celluloid clip D . The other end of the filter-paper strip led, through an M/10 potassium chloride solution and a saturated potassium chloride bridge, to a saturated calomel electrode. The electric arrangements used during the measurement of the potentials were the same as those described above.

The character of the stress in the exposed area could be instantly varied by turning the wire on its own axis. Thus when the exposed area occupied the outer arc of the bent specimen it was in tension; when it occupied the inner arc it was in compression. The influence of tensile and compressive stresses on the potential could be followed continuously by turning the specimen so as to bring the exposed area alternately into compression and into tension through a neutral position; in this way, complications caused by the general drift of potential with time were eliminated.

EXPERIMENTAL RESULTS

Series I. One-stage experiments with corrosive liquid applied up to fracture. The relation between stress range and lives of specimens is shown in figure 4, which indicates an obedience to the usual relation between the stress range S and the number of reversals before fracture, N , namely $N = Ce^{-kS}$, where C and k are constants.

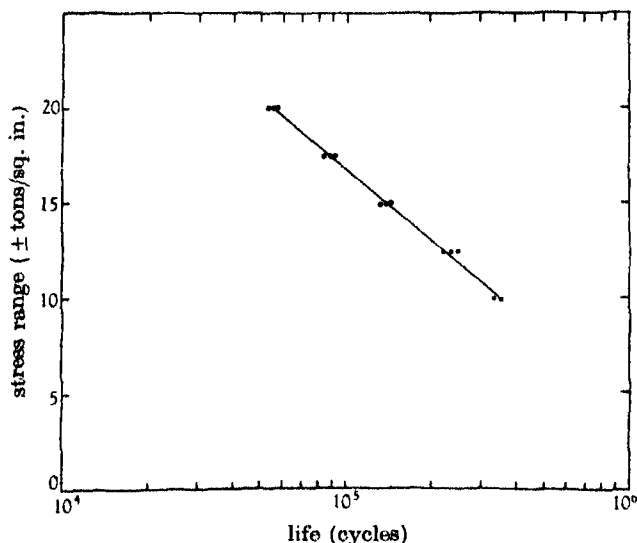


FIGURE 4. Corrosion-fatigue lives to fracture.

Series II. Two-stage experiments with corrosive liquid applied during the first stage only. In the main set of two-stage experiments, the specimens were dried at the end of the first stage, and exposed to alternating stresses in air during the second stage, the stress range being the same as in the first stage. Subsidiary experiments were carried out in which the specimen was wetted with a solution of potassium chromate, an efficient inhibitor, during the second stage. These gave practically the same results as when the second stage was conducted in air, indicating that any trace of chloride left behind after the washing was without influence, and also that the cooling action of a liquid was unimportant.

In figure 5 the total life (the sum of the two stages) is plotted against the corrosion period, both scales being logarithmic. It is clear that wetting with a corrosive liquid during the first stage for a time longer than a certain critical period very greatly reduces the total life, which, in the complete absence of corrosion, would be infinite, since the stress range always lies below the fatigue limit; on the other hand, the application of corrosive liquid for a still longer period may actually cause the total life once more to increase. The causes will appear later.

Series III. Measurement of iron passing into the combined state. The results shown in figure 6 make it clear that alternating stresses do increase not only the rate of mechanical damage but also the rate of chemical corrosion—a matter on which doubts had been expressed. The relationship between the stress range and the average rate of corrosion over a period of 60 min. is shown in figure 7; the rate of corrosion includes both the iron present as semi-adherent rust and that found in the liquid which has run off the specimen.

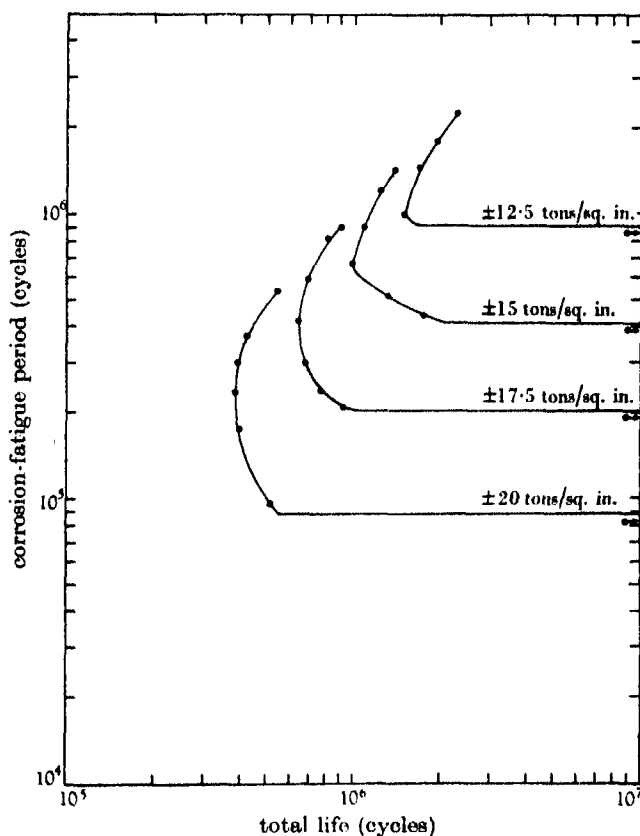


FIGURE 5. Two-stage corrosion-fatigue experiments.

Other experiments (figure 8) were carried out in $N/10$ hydrochloric acid. Here no insoluble corrosion product (rust) is formed. The rate of corrosion is constant until a fairly short time before breakage, when it becomes much faster. It is worthy of notice that, just at the final stage during which the presence of corrosive liquid hardly affects the rate of mechanical damage, the mechanical action greatly enhances the rate of corrosion.

In addition, experiments were carried out in which the corrosion product was collected at intervals of 60 min. from the single specimen. In figure 9 the results

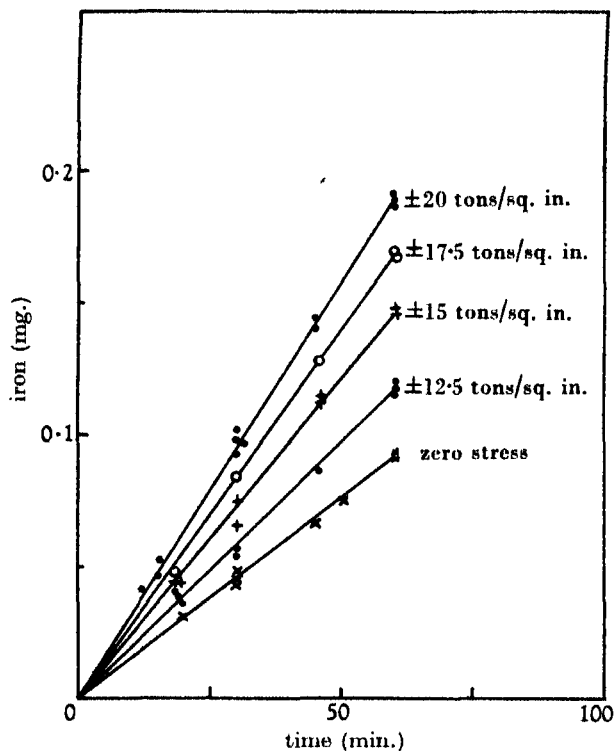


FIGURE 6. Initial rates of corrosion.

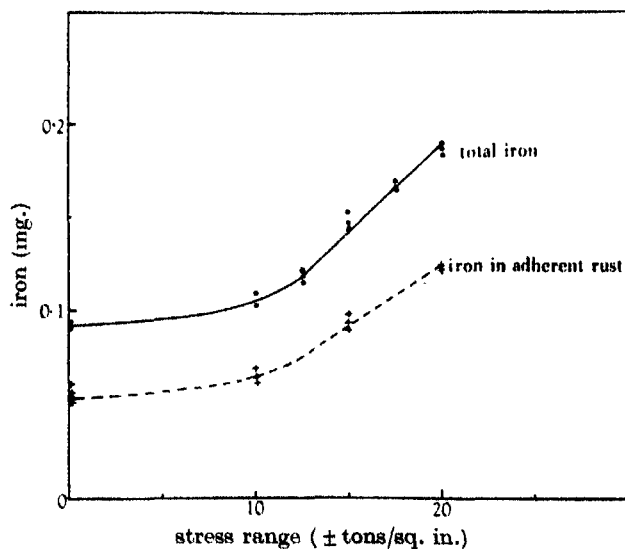


FIGURE 7. Influence of stress range on total iron corroded in 60 min.

are compared with the curve obtained from a number of specimens each subjected to corrosion fatigue for a different time, without interruption to rub off the rust. Comparison shows that the rubbing off of rust does appreciably increase the corrosion rate. Nevertheless the effect of alternating stress cannot be merely a loosening of the rust, since the quantity of rust remaining on the corroded specimen also increases with the stress range, as shown by the broken curve of figure 7.

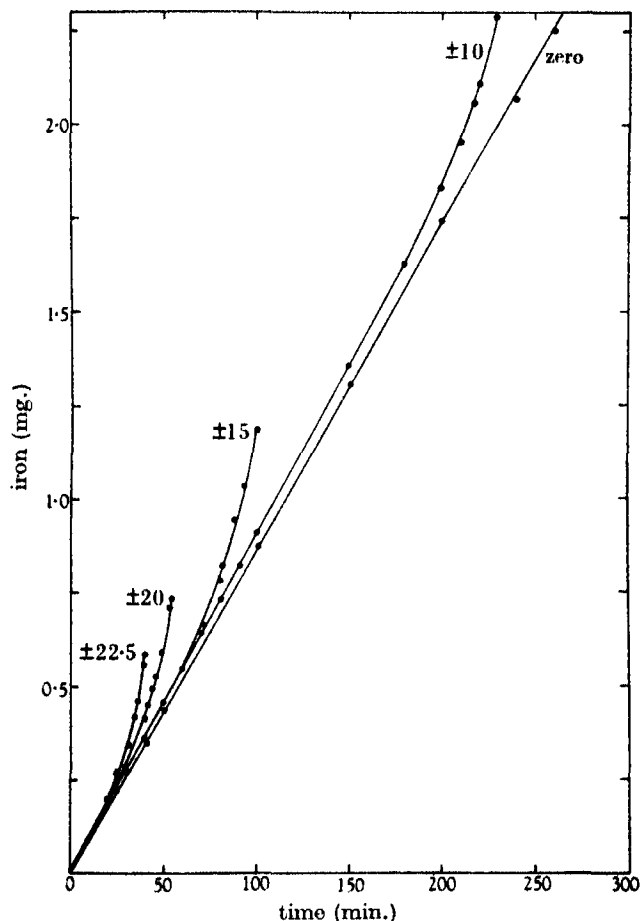


FIGURE 8. Rates of corrosion in N/10 hydrochloric acid.
Figures denote stress ranges in tons/sq.in.

Series IV. Experiments with cathodic current superimposed. The results shown in figure 10a indicate that small cathodic currents actually decrease the life of specimens, but, as the current is increased, the life is prolonged again. The number of pits and cracks is found to decrease steadily with increase in the applied cathodic current, whilst the quantity of iron compounds produced decreases also, as shown in figure 11. At a certain current, which increases with stress range, the life becomes

extremely long, the specimens being unbroken after 20 million cycles, and under these conditions no pits or cracks can be seen, and no iron in the combined state detected. It would seem, therefore, that a completely protective current does exist, at which corrosion fatigue is entirely arrested. With such a current superimposed, the corrosion-fatigue curve almost coincides with the air-fatigue curve (figure 10*b*). (Preliminary results with acid show that here a cathodic current, although diminishing corrosion, does not greatly prolong life.)

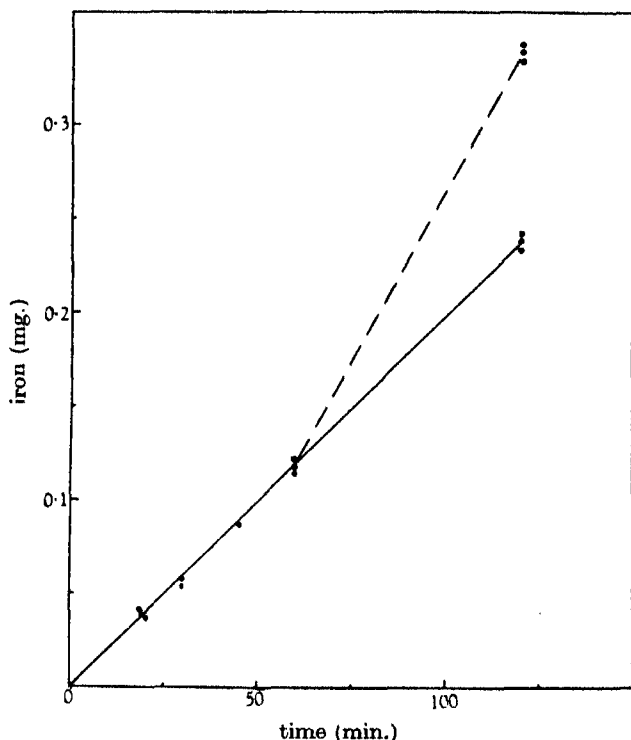


FIGURE 9. Influence of removal of rust on corrosion rate.

Series V. Measurements of movement of potential with time under alternating stress. The curves plotted in figure 12 show that there is always an initial rise in potential lasting a few minutes followed by a fall in potential which is more rapid at high stress ranges than at low ones. In figure 13 a logarithmic time-scale is employed so as to include the results of more lengthy experiments at low stress ranges. In all cases a final potential was reached which had roughly the same value for all ranges of stress. It is only whilst the potential is falling with time that the difference between the curves representing various stress ranges is so marked.

Series VI. Measurements of movement of potential with time under steady stress. Previous attempts to study the effect of stress on potential have been complicated by the fact that no two pieces of metal behave exactly alike, and that, quite apart from

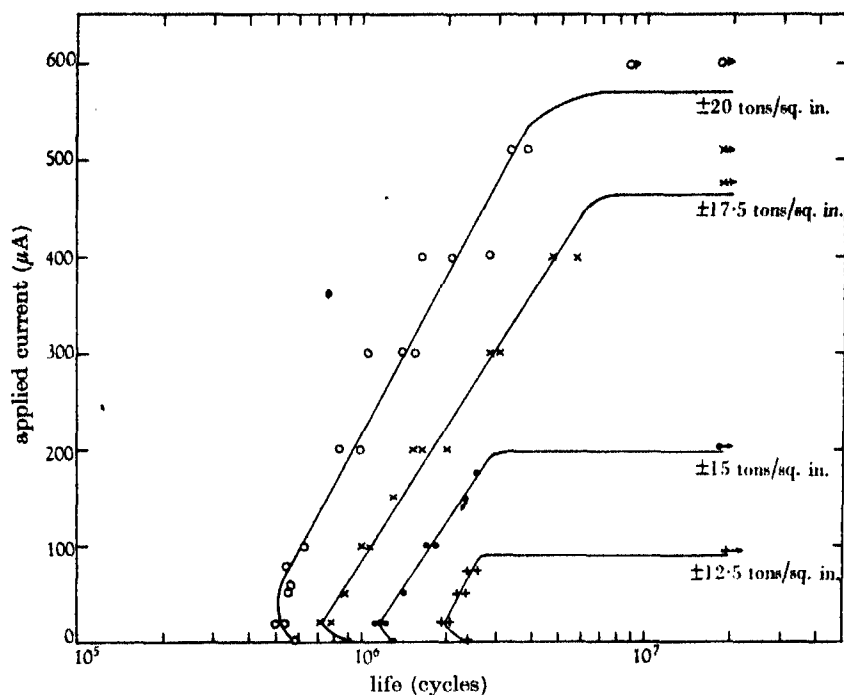


FIGURE 10a. Influence of applied cathodic currents on total life to fracture.

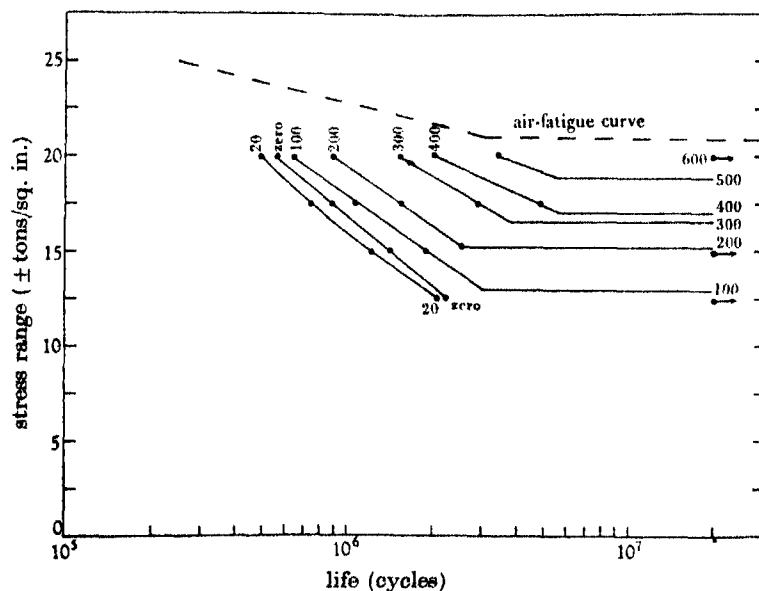


FIGURE 10b. Influence of stress range on life to fracture with applied currents. Figures denote applied currents in micro-amperes.

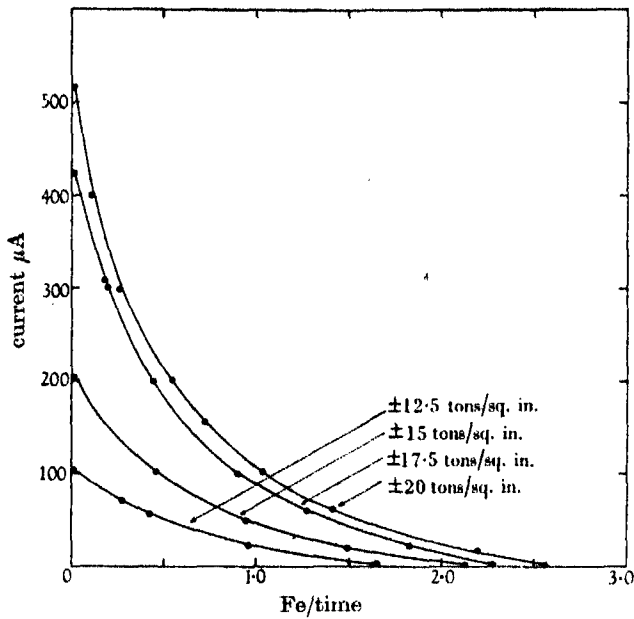


FIGURE 11. Average rates of corrosion with applied currents.

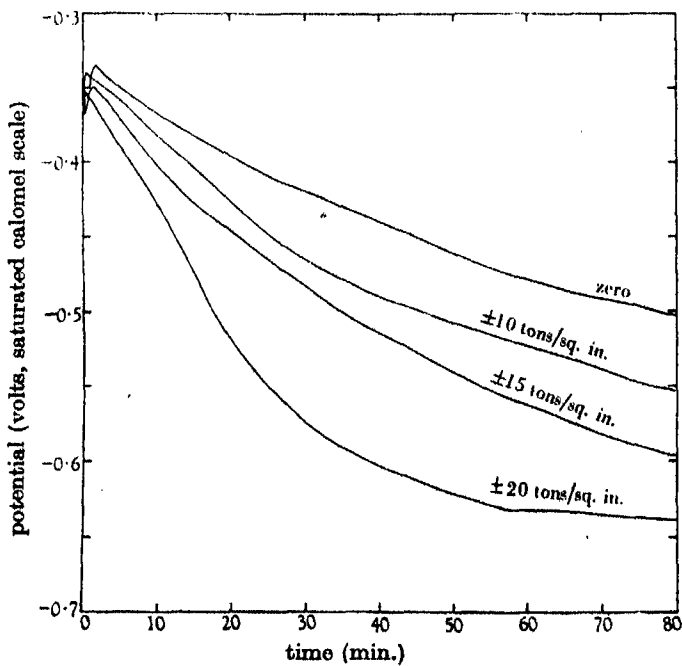


FIGURE 12. Influence of stress range on time-potential curves.

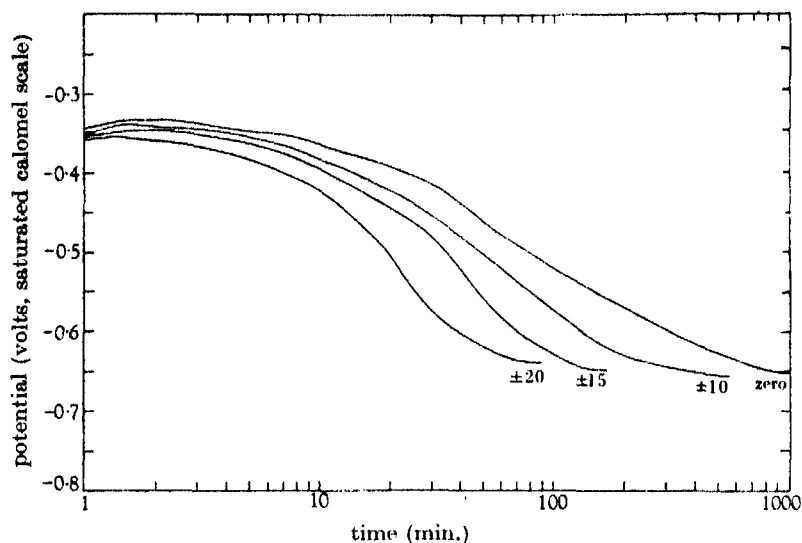


FIGURE 13. Time-potential curves to fracture. Figures denote stress range in tons/sq.in.

TABLE 2. INFLUENCE OF STATIC STRESS ON ELECTRODE POTENTIAL

time (min.)	potential of exposed area when in*					
	N_1	T	N_2	C	$N_1 - T$	$N_2 - C$
$\frac{1}{2}$	431	435	435	435	-004	000
1	438	438	438	438	000	000
2	456	456	457	456	000	001
4	484	478	484	484	006	000
7	—	—	510	507	—	012
10	507	504	513	506	003	007
	—	—	525	515	—	010
15	510	510	517	514	000	003
	—	—	548	526	—	022
20	526	526	529	525	000	004
	528	530	533	527	-002	006
	553	560	550	536	-007	014
25	545	548	547	540	-003	007
	576	573	568	556	003	012
	562	566	564	554	-004	010
30	565	569	567	555	-004	012
	586	586	583	570	000	013
	578	581	581	571	-003	010
35	574	574	571	551	000	020
	562	567	567	552	-005	015
40	587	591	590	583	-004	007
	589	594	594	588	-005	006
45	642	645	645	643	-003	002
	648	649	649	647	-001	002
50	672	675	675	675	-003	000
	680	680	680	680	000	000

* N , T , and C denote 'neutral', 'tension', and 'compression', respectively. The potentials are in millivolts on the saturated calomel scale.

applied stress, the potential drifts with time. The method described above allows the potential to be measured on a single specimen at comparable instants, so as to eliminate the effect of the general drift. The measurements recorded in table 2 show that at any time compressional stresses give higher (more noble) values than when the test area is in the neutral position; when the area is subjected to tension, the values tend to be, if anything, below those obtained without stress. As in the case of the tests under alternating stress, the same value is reached, whether the area is compressional, neutral or tensional; this same value was obtained in special experiments where there was no turning of the specimen.

INTERPRETATION OF RESULTS

Effect of corrosion period on life. The unexpected form of the curves in figure 5 may be explained by the shape of the pits. After short corrosion periods, these were rounded, whilst after longer periods they became elongated into cracks. Evidently the first effect of corrosion is to produce numerous pits, mostly hemispherical (as in the absence of stress). If one of these happens to be deeper than its neighbours, the stress concentration at the bottom will increase the e.m.f. between the anodic point (the stressed region at the bottom) and the cathode (the external surface of the specimen). The extra attack at the bottom will deepen and sharpen this particular pit, thus producing further stress intensification and further raising the e.m.f. Thus the pit, becoming steadily deeper and sharper, will penetrate in preference to its neighbours. However, as soon as the crack has become very sharp and deep, increase in resistance between anode and cathode will outweigh further increase in the e.m.f. Thus the rate of progress of the pioneer crack will fall off, and some of the neighbouring pits will in their turn develop into cracks.

If the corrosion stage is terminated at the period when isolated pioneer cracks have extended further than their neighbours, the stress concentration will already be considerable, and the subsequent life under dry fatigue will be short. When we reach the stage at which there are families of cracks instead of isolated cracks, the rapid increase of stress concentration will have ceased (Thum & Bruder 1938): there will then be only a gradual further shortening of the air-fatigue life.

It is now possible to explain the experimental facts. In figure 14, the corrosion-fatigue period is plotted against the logarithm of the life in the corresponding second stage in air. Provided that the first stage is shorter than a certain value (the 'incubation period'), the life endured in the second (dry) stage is very great. If the first stage exceeds that value, the duration of the second stage is extremely short; the incubation period probably represents the period during which the pioneer cracks are developing. If the corrosion stage is further prolonged, the shortening of the second stage proceeds more slowly than the corresponding lengthening of the first stage, so that the total life now increases with the increase of the corrosion period.

Certain relations, whose full significance cannot yet be assessed, were noticed between the incubation period T , the stress range S , and the total corrosion-fatigue life N as obtained in one-stage tests.

$$T = ae^{-bS}, \quad T = \alpha N^\beta,$$

T and N being expressed in cycles. The numerical values of the constants are

$$a = 5 \times 10^7, \quad b = 0.31; \quad \alpha = 48 \times 10^{-8}, \quad \beta = 2.45.$$

At each stress range, the minimum total life is approximately half the period endured when corrosion is continued to fracture.

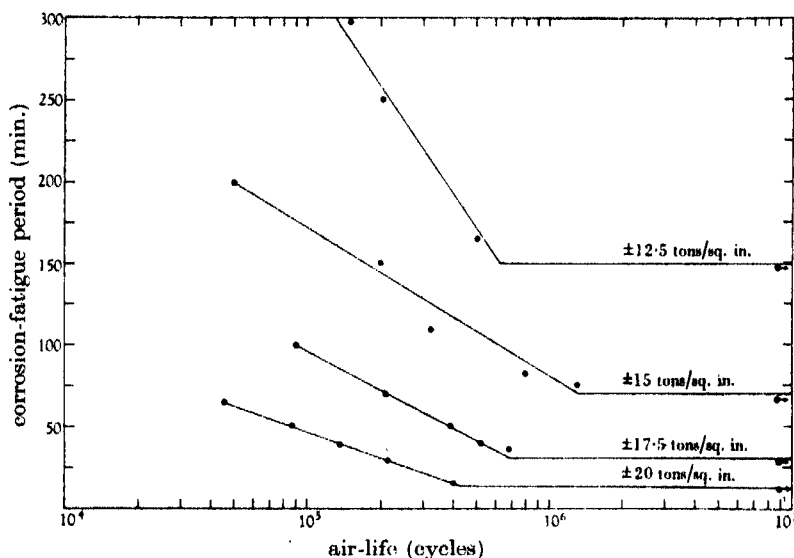


FIGURE 14. Relation between corrosion-fatigue periods and corresponding air-lives in second stage.

The results have potential importance to engineers—assuming that they extend to the lower frequencies common in service. In presence of alternating stresses unintermittent precautions must be taken to exclude corrosive influences. If vigilance is relaxed even for a short period, the final life of the stressed members may be shorter than if the precautions had not been resumed. The crack obtained may resemble a dry-fatigue crack, and will probably be erroneously attributed to dry fatigue.

Effect of cathodic protection on life. The fact that small cathodic currents actually decrease life, whilst stronger currents prolong it, can also be explained. The number of cracks is observed to diminish steadily as the cathodic current increases; the weakening effect of a few isolated cracks will be greater than that of a family of neighbouring cracks, owing to the greater stress intensification in the first case. Also the cathodic currents seem to diminish the number of cracks more rapidly than

they diminish the total corrosion, so that the intensity of attack will be increased, as in the case of a typical anodic inhibitor discussed elsewhere (Evans 1936; Evans & Chyzewski 1939).

The fact that corrosion is completely arrested by the application of a sufficient cathodic current density (figure 19) suggests a means of overcoming corrosion fatigue in absence of acid. Probably the cathodic protection can most easily be

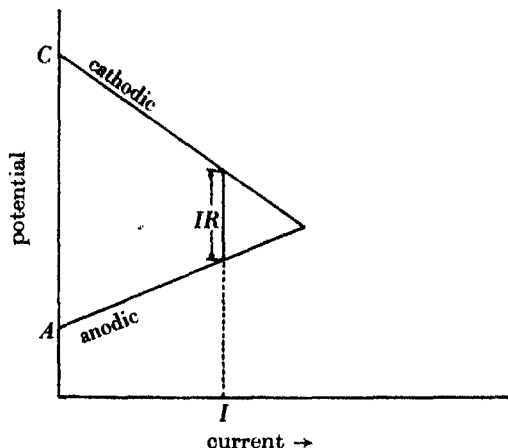


FIGURE 15

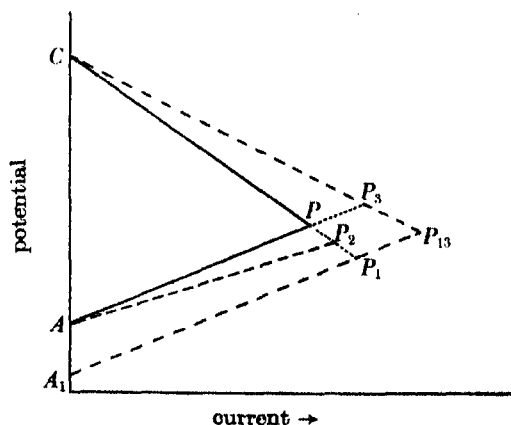


FIGURE 16

obtained by the application of special paints richly pigmented with metallic zinc (Mayne & Evans 1944; Huddle & Evans 1944). By rational application of electrochemical principles to the compounding of such paints, it is hoped to obtain better results than any yet achieved.

Mechanism of cathodic protection. In deciding between the three causes suggested for corrosion fatigue, the graphical method for corrosion velocity developed in this

laboratory will be used. The current flowing between anode and cathode (figure 15) is that value I which will produce an intercept IR between the anodic and cathodic polarization curves, where R is the resistance. The corrosion velocity will then be I/F , where F is Faraday's number. If R is assumed to be small, the abscissa of the intersection can be taken to represent the corrosion current, and at the outset this assumption will be made.

If (figure 16) the anodic and cathodic curves, in absence of stress, are AP and CP , the corrosion current will be the abscissa of P . If, as in mechanism 1 (p. 373), the stresses are considered to shift the anodic curve downwards, say to A_1 , the current will be increased to the abscissa of P_1 . If, as in mechanism 2, the effect of stress is to diminish anodic polarization, making the anodic curve less steep (AP_2), the current will become the abscissa of P_2 . If, as in mechanism 3, the anodic curve is left at AP and cathodic polarization is reduced, giving a new cathodic curve CP_3 , then the enhanced current will be the abscissa of P_3 . Either type of *anodic* stimulation *depresses* the potential corresponding to the intersection, whilst a decrease of *cathodic* polarization will *raise* it.

The effect of excess cathodic current from an external anode is shown in figure 17, based upon the principles developed by Hoar (1938) and by Mears & Brown (1938). If the applied cathodic current is represented by the length YZ , the corrosion current will be reduced from WP to XY (neglecting resistance and assuming that cathodic current does not alter distribution of anodic and cathodic areas). A current sufficiently strong to depress the potential to the value of an unpolarized anode, A , will prevent corrosion altogether.

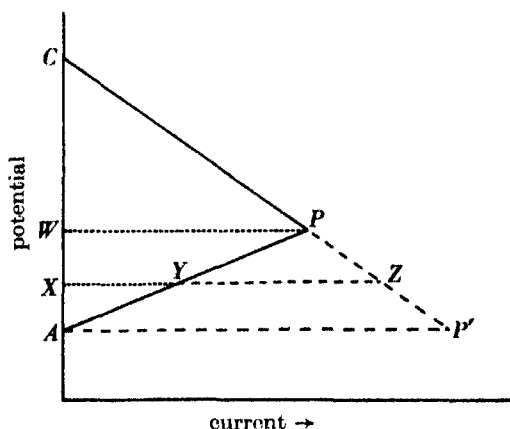


FIGURE 17

The experiments without cathodic protection showed that corrosion velocity increases with increase of stress range. Thus the corrosion current must have increased, and yet the potential was practically unaltered. This is best explained by assuming that mechanism 3 is acting along with mechanisms 2 (*b*), 2 (*a*) or 1. Since

the number of cracks per unit length was found to increase with stress range, the anodic polarization must have been diminished; also the resistance must have been diminished, so that, notwithstanding the increased current, the intercept should have been brought closer to the intersection. Thus mechanisms 2 (a) and 2 (b) can

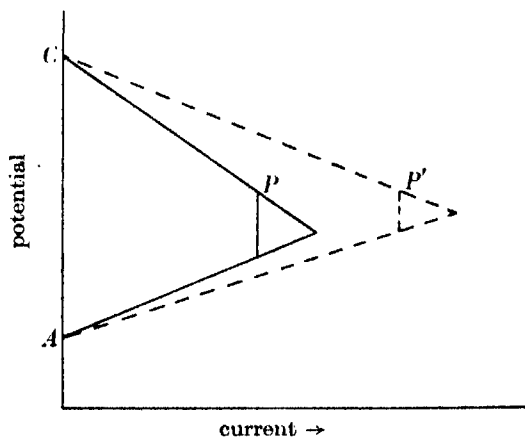


FIGURE 18

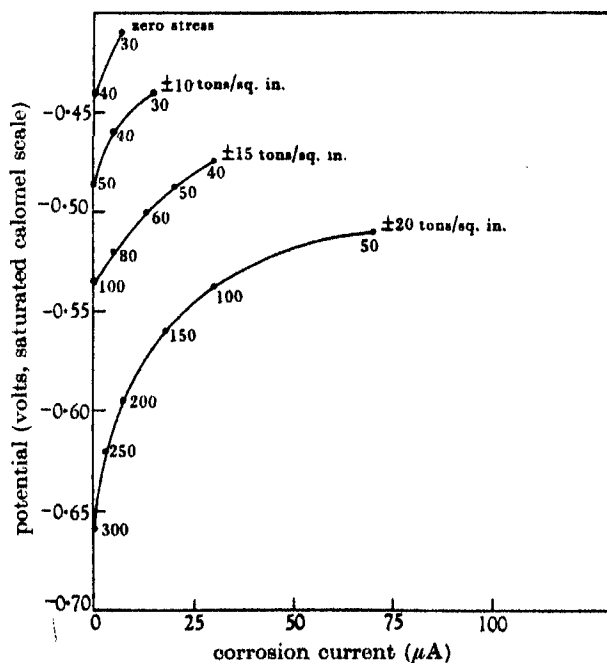


FIGURE 19. Relation between potential and corrosion currents as calculated from corrosion rates with applied currents. Figures at individual points denote applied currents in microamperes.

be assumed, without *ad hoc* assumptions, to operate. It is not absolutely necessary to introduce mechanism 1 in order to explain the results.

Figure 18 suggests the changes brought about by increased stress. The current is increased from a value represented by the abscissa of P to that represented by the abscissa of P' , the potential remaining much the same.

The study of the fall of potential with time under alternating stress (series V) or under static stress, whether compressional or tensional (series VI), led to the same final value irrespective of stress range, or of the nature of the applied stress. This suggests that below the elastic limit, the effect of stress on the potential of metal not covered with a quasi-protective film is negligible. A stress which merely produces elastic deformation in metal may of course rupture a film, or open more widely any discontinuities already existing; this doubtless explains why the fall is more rapid at high ranges of stress than low ones, and why compressional stresses (which should close up apertures) tend to raise the potential.

It is probable, but not definitely proven, that stress in the plastic range (or residual stresses due to previous straining) does affect the potential.

We would like to thank Dr J. N. Agar, Dr A. J. Gould, Dr T. P. Hoar, Dr J. E. O. Mayne and Mr A. U. Huddle for stimulating discussion; thanks are also due to the Corrosion Committee of the Iron and Steel Institute, and especially Mr W. A. D. Forbes, for assistance in connexion with apparatus. One of us (M.T.S.) would acknowledge his indebtedness to the British Council for the grant of a scholarship.

REFERENCES

- Bacon, F. 1935 *Trans. N.-E. Cst Instn Engrs Shipb.* **51**, 216.
Behrens, O. 1933 *Mitt. Wöhler-Inst.* Heft 15.
Bengough, G. D. 1932 *J. Inst. Met.* **49**, 93.
Evans, U. R. 1936 *Trans. Electrochem. Soc.* **69**, 213.
Evans, U. R. & Chyżewski, E. 1939 *Trans. Electrochem. Soc.* **76**, 215.
Fink, C. G., Turner, W. T. & Paul, G. T. 1943 *Trans. Electrochem. Soc.* **83**, 377.
Gerard, I. J. & Sutton, H. 1935 *J. Inst. Met.* **56**, 33.
Gill, E. T. & Goodacre, R. 1936 *Carnegie Schol. Mem.* **35**, 139.
Godfrey, H. J. 1941 *Trans. Amer. Soc. Metals*, **29**, 133.
Gough, H. J. 1932 *J. Inst. Met.* **49**, 1.
Gough, H. J. 1932 *J. Inst. Met.* **49**, 93.
Gough, H. J. 1935 *Proc. Staffs Iron Steel Inst.* **50**, 74.
Gough, H. J. & Sopwith, D. G. 1932 *Proc. Roy. Soc. A*, **135**, 392.
Gough, H. J. & Sopwith, D. G. 1934 *J. Inst. Met.* **54**, 193.
Gould, A. J. 1934 *Engineering*, **138**, 79.
Gould, A. J. 1936 *Engineering*, 8 May.
Gould, A. J. & Evans, U. R. 1939 *2nd Rep. Alloy Steels Res. Comm.* p. 325.
Haigh, B. P. 1917 *J. Inst. Met.* **18**, 55.
Haigh, B. P. 1929 *Trans. Instn Chem. Engrs*, **7**, 29.
Hoar, T. P. 1938 *J. Electrol. Depos. Soc.* **14**, 33.
Huddle, A. U. & Evans, U. R. 1944 *J. Iron Steel Inst.* **149**, 109 P.
Krystof, J. 1935 *Metallwirtschaft*, **14**, 305.
Mayne, J. E. O. & Evans, U. R. 1944 *Chem. Ind.* p. 109.

- McAdam, D. J. 1926-1931 *Proc. Amer. Soc. Test. Mater.* (II) **26**, 224; **27**, 102; **29**, 250; **30**, 411; **31**, 259.
- McAdam, D. J. & Clyne, R. W. 1940 *Bur. Stand. J. Res., Wash.*, **29**, 685.
- Mears, R. B. & Brown, R. H. 1938 *Trans. Electrochem. Soc.* **74**, 519.
- Shelton, S. M. & Swanger, W. H. 1935 *Bur. Stand. J. Res., Wash.*, **24**, 17.
- Speller, F. N., McCorkle, I. B. & Mumma, P. F. 1928 *Proc. Amer. Soc. Test. Mater.* **28**, 159.
- Stuart, N. & Evans, U. R. 1943 *J. Inst. Met.* **147**, 131 P.
- Thum, A. B. & Bruder, W. 1938 *Dtsch. Kraftfahrtforschung*, No. 11.
- Whiteley, J. H. & Hallimond, A. F. 1925 *Carnegie Schol. Mem.* **14**, 163.

The electric conductivity and the activation energy of ionic migration of molten salts and their mixtures

BY H. BLOOM AND E. HEYMANN

Chemistry Department, University of Melbourne, Melbourne, Australia

(Communicated by Sir David Rivett, F.R.S.—Received 4 March 1946)

The electric conductivity (κ), and its variation with temperature, of many molten salts of predominantly ionic character can be represented by a simple exponential equation $\kappa = \text{const.} \times e^{-C/RT}$. Deviations from this relation are sometimes found for partially covalent compounds (e.g. ZnCl_2 , PbCl_2) where constitutional changes may be expected with change of temperature. The activation energy of ionic migration (C) is always smaller than the activation energy of viscous flow. This fact is attributed to the difference in the configurational changes that occur in the two processes. For alkali chlorides, C decreases with increasing ratio of anion to cation radius. For electrolytes involving multivalent ions, C is greater than for uni-univalent ones. Increasing amount of covalency of the bonds involved tends to lower C .

The conductivities of a number of mixtures of electrolytes (CdCl_2 - CdBr_2 , CdCl_2 - PbCl_2 , CdCl_2 - NaCl , CdCl_2 - KCl , PbCl_2 - KCl) were measured over a range of compositions and temperatures. The activation energies of ionic migration and, where possible, the equivalent conductivities were calculated, and the results discussed together with those obtained in other systems by various investigators.

In no system so far investigated is the conductivity a linear function of the composition expressed as mole fraction. In systems which give no evidence of complex ion formation in the mixture, the conductivity usually shows moderate negative deviations from additivity (e.g. CdCl_2 - CdBr_2). Only one system so far shows a positive deviation from additivity (CdCl_2 - PbCl_2).

Strong negative deviations from additivity are found in systems in which complex ions are likely to exist in the mixtures (PbCl_2 - KCl , CdCl_2 - KCl , CdCl_2 - NaCl). In the systems CdCl_2 - KCl and PbCl_2 - KCl , the conductivity isotherms have minima at all temperatures investigated; in these systems, the phase diagram indicates a congruently melting compound. Additional minima in the conductivity isotherms are found near compositions at which the phase diagram indicates incongruently melting compounds, but only at low temperatures; at higher temperatures, these minima disappear.

The activation energies (C) have maximum values near compositions that correspond to unstable compounds; in this case, C contains part of the energy change involved in the transition from the complex to the simple ions.

In some cases, the activation energy (C) of the molten systems rises to very high values as the crystallization temperature is approached. This is interpreted as being due to the existence of a high degree of order in the melt just above the melting-point.

Further relations between conductivity, and its temperature coefficient, the activation energy of ionic migration and the constitution of the molten salt mixtures are discussed.

1. INTRODUCTION

Biltz & Klemm (1926) gave a comprehensive discussion of the electric conductivity of molten halides, mainly with reference to the position of the cationic element in the periodic table and the degree of dissociation. Recently, Mulcahy & Heymann (1943) have discussed the conductivity and constitution of molten salts and their mixtures in the light of further experimental evidence and modern theory.

Most of the previous investigators of mixtures, e.g. Sandonnini (1920), report conductivity isotherms only at one temperature. However, for a proper understanding of mixtures, particularly in order to ascertain whether more or less stable complex ions are present, it is essential to obtain information also about the temperature coefficients and the activation energies of ionic migration. Investigations over a range of temperatures were carried out in systems involving calcium chloride, magnesium chloride and alkali chlorides by Karpachev & Stromberg (1934, 1935), Barzakowski (1940), Klocho (1940) and Shcherbakov & Markov (1939). Because of the high temperatures involved, conductivity cells (of a capillary type), made from silica, had to be used. However, it is doubtful whether results of great accuracy can be obtained in these systems because molten alkali chlorides attack silica, and, as a consequence, the cell constant may alter during a series of experiments.

It was our aim in this investigation to study further systems over a range of temperatures, particularly those in which complex formation could be expected in the melt. We confined ourselves to mixtures which are liquid in a sufficient range of temperatures below 740° C. Such systems can be investigated in cells made from Supremax glass (Jena). This glass is not seriously attacked even by alkali chlorides. It softens above 800° C and shows no deformation below 750° C.

2. THE ACTIVATION ENERGY OF IONIC MIGRATION OF MOLTEN SALTS

In order to understand the mechanism of electric conductivity of molten salts, it is useful to discuss the solid salts first. The molten system may then be treated as a very disordered solid, making use of concepts due to Frenkel, Schottky, Jost and others (cf. Mott & Gurney 1940). These investigators have related diffusion and ionic conductance in solid salts with irreversible and reversible imperfections of the crystal lattice.

The high temperature, i.e. structure-insensitive, conductivity of a solid salt is proportional to the number of lattice defects per c.c. and to their mobility. Both factors vary exponentially with temperature, and an expression

$$\kappa = \text{const.} \times e^{-\frac{1}{2} \frac{W+U}{kT}} \quad (1)$$

is obtained, where W is the energy necessary to produce the lattice defect and U is the height of the energy barrier involved in its migration. Strictly speaking, equation (1) applies to the case in which conductivity is due to the migration of one type of defect, i.e. to that of cation or anion alone (Mott & Gurney 1940).

In the molten state, the lattice is presumably disordered to such an extent that the fraction of mobile ions may be assumed to be very great and, within certain limits, virtually independent of temperature, provided no constitutional changes take place with change of temperature. Hence the contribution of each ionic species to the conductivity will be proportional to an exponential term containing only the U -term of equation (1). For a purely ionic binary melt, a relation of the form

$$\kappa = A_1 e^{-C_1/RT} + A_2 e^{-C_2/RT}, \quad (2)$$

may be expected, where C_1 and C_2 may be called the activation energies of ionic migration of cation and anion respectively, and κ is the specific conductivity. If the activation energies of cation and anion are nearly equal, an equation

$$\kappa = A e^{-C/RT} \quad (3)$$

may be expected to fit the experimental results. Also, within the limited temperature intervals that are available for experiments, equation (3) will often hold approximately if the activation energies of cation and anion are very different because, in that case, the species with the higher activation energy may contribute very little to the conductivity, or, in other words, conductance will be due predominantly to cation or anion alone.

The experimental data, particularly at high temperatures, may not be of sufficient accuracy to provide a reliable test for equation (3). Nevertheless, for ionic melts such as alkali and alkaline earth halides, $\log \kappa$ gives an approximately linear plot against $1/T$ in the experimental range, and the same is found for AgCl , TiCl , PbBr_2 and CdCl_2 . On the other hand, equation (3) does not satisfactorily describe the variation of conductivity with temperature for AgBr , LaCl_3 , PbCl_2 and ZnCl_2 , $\log \kappa$ against $1/T$ showing non-linear plots. In the case of zinc chloride and lead chloride, the bonds involved are of a more covalent character than, for instance, with the alkali halides. With these salts (ZnCl_2 , PbCl_2), the breakdown of equation (3) may be due to constitutional changes which may occur on variation of temperature. Such processes may change the number of ions as well as their transference numbers and may thus affect A and C in equation (3).

The activation energy may also be represented as C' , as defined by the equation

$$A = A' e^{-C'/RT}, \quad (3a)$$

where A is the equivalent conductivity. The latter is defined as

$$A = \frac{\kappa E}{d}, \quad (4)$$

where E is the equivalent weight and d the density. If the activation energies of ionic migration of various salts are compared, the use of C' is preferable to that of C , because C' refers to a state in which we have always one equivalent of the salt between electrodes at 1 cm. distance. On the other hand, for a comparison of the activation energy of ionic migration with that of viscosity (vide below), it is prefer-

able to use C , which is calculated from κ (equation (3)) because the viscosity is dimensionally more closely related to κ than to A .

Theoretical and experimental work, in particular that by Dunn (1926), Andrade (1934), Eyring (1936), Ward (1937) and Barrer (1943), has shown that the viscosity of liquids, including ionic liquids, may be represented by an equation similar to equation (3), namely

$$\eta = \text{const.} \times e^{B/RT}, \quad (5)$$

where B is the activation energy of viscosity.

In table 1, values of the activation energy of ionic migration (C, C') and of the activation energy of viscosity (B), both in kcal./mol., are shown. Some values for B are taken from Barrer's paper; others are calculated from viscosity data by Karpachev & Stromberg (1938). Most values for A, C, C' and λ are calculated from conductivity and density data quoted by Biltz & Klemm; others are computed from our conductivity values for PbCl_2 , CdCl_2 , CdBr_2 (cf. §§ 4, 5) and LiBr .* The ratio of anion to cation radius (r_A/r_C) is that of the crystal radii calculated by Pauling (1940). The λ -values are those at a temperature 10 % above the melting-point (abs.). The choice of a temperature related to the melting-point (rather than the boiling-point) as a corresponding temperature, is indicated when the melt is regarded as a disordered lattice. The melting-point itself has not been chosen because of the possibility of incipient lattice formation immediately above the melting-point (cf. §§ 7-9).

While the absolute values of C and C' often differ considerably, the difference being greater with the alkali halides than with all other types of salts, the order of the salts in the various groups is the same. The experimental accuracy of C' is somewhat less than that of C because C' contains the experimental errors of the density determination in addition to those of the conductivity determination.

It is not certain whether the values of viscosity of molten salts are sufficiently accurate to permit a comparison between the B -values within the various series of salts in table 1. However, as far as the order of magnitude is concerned, it can be seen that in all cases the activation energy of viscosity (B) is much greater than that of ionic migration (C), the ratio B/C being about 3-6 for uni-univalent electrolytes and about 2 for bi-univalent electrolytes. We suggest that this great numerical difference between B and C is connected with the difference in mechanism between conductivity and viscous flow in molten salts. Viscous flow involves a breakdown of the configuration of both anions and cations and may be expected to depend to a much greater extent on the larger sized anions than on the smaller sized cations

* LiBr . The salt used was a 'Merck puriss' product which was found to contain 99.9 % LiBr . The conductivity was measured (between 550° and 610° C) as described in § 3, and nitrogen was passed through the apparatus in order to avoid oxidation and hydrolysis. The data may be expressed by the equation

$$\kappa = 4.63 + 0.0056(t - 550),$$

where t is in °C. From the linear plot of $\log \kappa$ against $1/T$, we obtain $C = 1.75$ kcal./mol.

(Frenkel 1937). On the other hand, in electric conductance through molten alkali chlorides, the contribution of the small cations will be much greater than that of the larger anions. In the case of the chlorides of many bivalent metals, conductance will be due mainly to the ions which have the lower charge.

TABLE 1.

	<i>A</i>	<i>C</i>	<i>C'</i>	<i>B</i>	r_A/r_0	<i>A</i>
LiCl	183	1.15	1.72	8.8	3.01	11.5
NaCl	150	1.54	2.70	9.4	1.91	7.3
KCl	120	2.30	3.26	7.8	1.36	6.5
RbCl	94	2.83	3.66	—	1.22	6.5
CsCl	86	3.33	3.75	—	1.07	6.6
LiBr	~ 177	1.75	—	6.0(?)	3.25	13.5
NaBr	148	1.84	2.58	10.6	2.05	7.4
KBr	109	2.55	3.42	7.9	1.47	5.8
NaI	150	1.25	2.00	7.4	2.27	4.8
KI	104	2.75	3.70	9.2	1.62	5.5
AgCl	118	0.99	1.20	5.3	1.44	7.4
AgBr	99	0.70–1.00	0.99	4.5–5.4	1.55	—
AgI	105	0.80	0.90	5.8	1.71	4.0
CuCl	94	0.85	1.06	5.5	1.89	6.2
MgCl ₂	35	3.56	3.97	—	—	6.4
CaCl ₂	64	4.10	4.72	9.5	—	13.5
SrCl ₂	69	4.00	4.65	—	—	11.0
BaCl ₂	77	4.15	4.35	—	—	10.5
CdCl ₂	58.5	2.30	2.13	4.5	—	7.3
CdBr ₂ *	41.2	2.75	—	—	—	5.4
PbCl ₂ †	53.0	3.30–4.50	4.20	6.6	—	—
PbBr ₂	27.0	4.35	4.32	9.2	—	16.0

* The density of CdBr₂ is 3.99 at 651° C (unpublished results by F. H. Dorman in this laboratory).

† The *B*-value of PbCl₂ has been calculated from determinations of viscosity by B. S. Harrop in this laboratory (not yet published).

The assumption that molten alkali halides are predominantly, though not exclusively, cation conductors receives support on comparing the equivalent conductivities (*A*) of molten alkali halides (table 1). It is seen that a change of cation involves a large change of *A*, whereas the change of *A* on changing the anion is much smaller. However, the contribution of the anions to the conductance is not likely to be zero because, even in the solid state, above 400–500° C, alkali halides have transference numbers of the anion which are small but not zero. The silver halides (except AgI), which are cation conductors in the solid state up to the melting-point (Tubandt, Reinhold & Liebold 1931), will probably show a similar type of conductance in the molten state, a conclusion which receives some support by the fact that *A* and *C* do not vary strongly with change of anion.

The alkaline earth chlorides and the chlorides and bromides of lead and cadmium are anion conductors in the solid state. In the molten state, there is only moderate

variation of A , C and C' as we go from calcium to barium. In the case of the halides of lead and cadmium, both A and C vary strongly on change of anion. The fact that CdCl_2 , MgCl_2 and PbCl_2 have somewhat smaller A -values than the alkaline earth chlorides runs parallel to their partially covalent character. The data are not sufficient to prove that these molten electrolytes are predominantly anion conductors, but they are compatible with this assumption.

Table 1 shows that B/C is smaller for electrolytes with bivalent cations than for uni-univalent ones. This fact may be due to a greater similarity of mechanism between viscosity and conductivity of electrolytes with a bivalent cation, the anions in this case probably dominating both in conductance and viscous flow.

It is thus seen that, *with many molten salts, viscous flow involves a different configurational change and higher energy barriers than ionic migration*. This may be the reason for the absence of a simple relation between conductivity and viscosity of molten salts. For instance, the temperature coefficients of the two properties are not the same. Moreover, in mixtures of molten salts, the variation of conductivity and viscosity with composition does not show even qualitatively the relation that an increase of conductivity corresponds to a decrease of viscosity and vice versa (Barzakowski 1940; cf. also Mulcahy & Heymann 1943).

The predominance of ionic migration of one ionic species in the conductance of molten salts must be due to a great difference between the activation energies of cation and anion. This is probably the reason why the conductivity of most simple ionic melts can, in a limited range of temperature, be represented by equation (3).

It is seen from table 1 that, for alkali and alkaline earth halides, high activation energies (C and C') generally correspond to low values of A and vice versa. Among the alkali halides there is a distinct increase of C and C' and a decrease of A , from lithium to caesium, together with a decrease of the ratio of anion to cation radius (r_A/r_C) and a simultaneous increase of the amount of co-ordination. The relative smallness of the activation energy (C and C') of lithium halides and of NaBr is probably due to low energy barriers in a system in which r_A/r_C is great. In the case of the lithium halides, the small cations must have considerable freedom of movement in the interstices between the large anions in contact.

With the alkaline earth chlorides, the values for C and C' are higher than those of the alkali chlorides, first because of the stronger electric field in an assembly containing bivalent ions, and secondly because these salts are likely to be predominantly anion conductors. Anions because of their larger relative size may be expected to possess larger activation energies than cations of the same valency.

An inspection of table 1 suggests also that, for salts of the same valency type, the activation energies of ionic migration are smaller when the salts involved are of a more strongly covalent character than when they are predominantly ionic. For instance, MgCl_2 and CdCl_2 have smaller activation energies than the alkaline earth chlorides, and the silver halides and CuCl have very much smaller activation energies than the alkali halides. It is suggested that the low activation energies in melts

of a more covalent bond character are connected with a 'smearing out' of energy barriers due to ion polarization.

Only brief reference is made to A , the temperature-independent term of equation (3), because its theoretical significance is much less easy to interpret than that of C ;

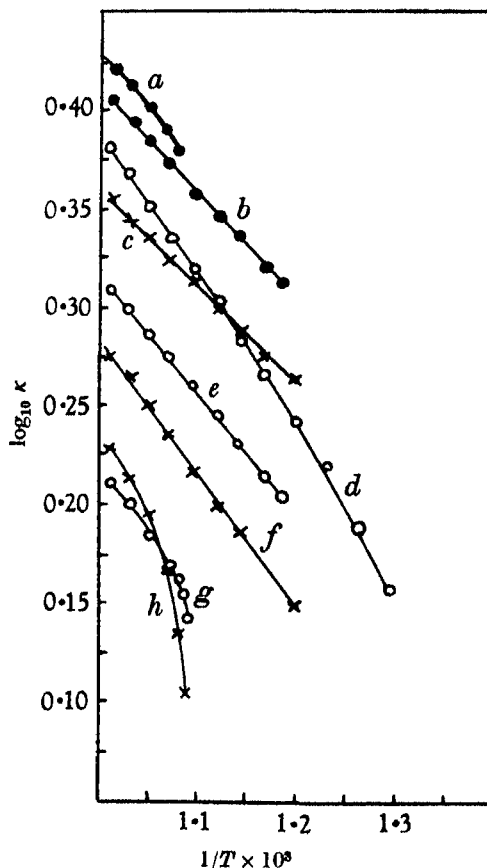


FIGURE 1. Plot of $\log \kappa$ against $1/T$ for a number of molten salts and mixtures. (a) 26.0 mol. % CdCl_2 + 74.0 mol. % NaCl ; (b) 34.1 mol. % CdCl_2 + 65.9 mol. % NaCl ; (c) CdCl_2 ; (d) PbCl_2 ; (e) 79.7 mol. % CdCl_2 + 20.3 mol. % KCl ; (f) 63.8 mol. % PbCl_2 + 36.2 mol. % KCl ; (g) 20.4 mol. % CdCl_2 + 79.6 mol. % KCl ; (h) 20.5 mol. % PbCl_2 + 79.5 mol. % KCl .

and also because its calculation from the experimental data may involve considerable inaccuracy, a small error in the activation energy producing a large error in A .^{*} There is comparatively little variation of A for similar salts such as alkali halides (except lithium salts). Moreover, the silver halides, as well as CuCl , have values of A not very different from those of the alkali chlorides. The variation of A for the various compounds shows little relation to that of the temperature-independent

* The dimension of A (table 1) is that of specific conductivity.

term of the viscosity equation (equation (5)), as calculated by Ward and Barrer. These points will be discussed in a later publication on the viscosity of molten salts and their mixtures.

Figure 1 shows a number of characteristic curves of $\log \kappa$ against $1/T$ for several simple salts and mixtures, based on experimental results of §§ 5-8.

3. EXPERIMENTAL

Owing to the very high conductivity of molten salts, it is necessary to use a conductivity cell with a high cell constant. Only cells involving either a constricted U-tube or a capillary can therefore be expected to give results of sufficient accuracy. A constricted U-tube has certain disadvantages, and a cell of the capillary type, shown in figure 2, similar to those used by Lorenz & Kalmus (1907) and Aten (1910) was therefore used. One Pt-electrode was inside the Supremax glass tube, which

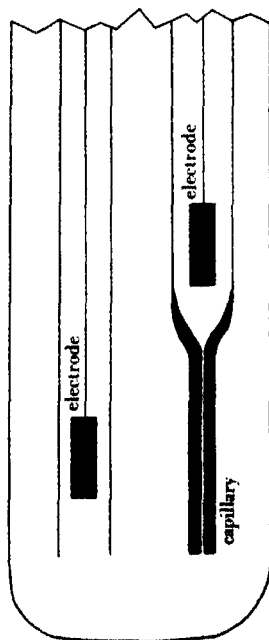


FIGURE 2. Conductivity cell.

ends in the capillary, and the other Pt-electrode was outside. This arrangement permits stirring of the melt, thus facilitating the attainment of uniform temperature. Slight variations in the position of the electrodes hardly affect the cell constant, which is mainly determined by the dimensions of the capillary. However, an arrangement as shown in figure 2 can give reliable results only when the conductivity through the glass wall of the capillary tube is negligible compared with that of the melt.

By measuring the resistance of our cell after the capillary was sealed at the end and comparing this value with that obtained in the usual arrangement, with the open capillary, we satisfied ourselves that, even at the highest experimental temperature (720°C), the resistance with the capillary closed was more than 175 times greater than with the open capillary. At 700°C the resistance through the glass wall was even higher. Hence up to 720°C the conductance through Supremax glass can be neglected.

In most experiments, the container carrying the bulk of the melt was made from Moncrieff combustion glass, which remains undeformed up to 740°C .

The resistance was measured by a Wien bridge circuit, using Tinsley induction-free shielded resistances and a Sullivan inductance-free triple ratio arm. Variable condensers were arranged parallel with the resistances. A valve oscillator was used as a source of current. The point of balance, determined by a telephone, was not affected by the frequency used; in most experiments, the frequency was about 3000 cyc./sec. The shielded bridge circuit was brought to earth potential by a Wagner earth. The measurements of resistance could be made to within 0.1 to 0.2 %.

The temperature of the melt, measured by a Pt-PtRh (10 %) thermocouple and a Leeds and Northrup potentiometer, could be determined accurately to $\frac{1}{2}^{\circ}\text{C}$. The well lagged furnace contained two stainless steel tubes, welded at right angles at their centres. They were wound with nichrome wire, the horizontal and vertical tubes being heated by separate circuits. Before recording any reading, the temperature was constant for 5 to 10 min., and we satisfied ourselves that the temperatures at different levels were constant within 0.5 to 1°C . As the temperature coefficient of conductivity of molten salts is of the order of 0.1 to 0.2 % per degree, the limit of accuracy due to variation of temperature is of that order.

The cell constant was determined by means of molten lead chloride using the conductivity values by Lorenz & Kalmus (1907). It was also determined independently, using a N-KCl solution, and the value thus obtained was corrected for expansion of the glass.* The two methods give values that agree within 0.3 %.

The cell constant was checked at frequent intervals and was found not to change appreciably during prolonged use of a capillary. One capillary, for instance, after use with twelve molten salt mixtures had changed its cell constant by less than 0.3 %.

The salts used in this investigation were either of A.R. purity, which was checked by analysis, or prepared by suitable methods. Cadmium bromide was made from a 'May and Baker, Pure' product by fusing while a stream of HBr was passed through. Cadmium chloride was made from pure electrolytic cadmium (Electrolytic Zinc Co., Tasmania), dehydrated and fused while a current of dry HCl was passed through. Standard methods were employed for analysis. The electro-analytical method was used for cadmium.

* The coefficient of expansion of Supremax glass is given by Espe & Knoll (1936) as 33×10^{-7} . Its composition is stated to be (in per cent) 56.4 SiO_2 , 8.87 B_2O_3 , 20.13 Al_2O_3 , 4.8 CaO , 0.63 Na_2O , 0.64 K_2O , 8.65 MgO .

4. THE SYSTEM CADMIUM CHLORIDE-CADMIUM BROMIDE

The phase diagram (Nacken 1907) indicates solid solutions over the whole range. There is no evidence of intermediate phases.

Figure 3 shows the variation of the specific conductivity (κ) and of its temperature coefficient $d\kappa/dT$ with composition (molar fraction). There is a slight negative deviation from additivity in the conductivity isotherms, and the temperature coefficient shows a gradual change without a turning point. Preliminary density determinations by F. H. Dorman in this laboratory indicate that the density of the mixtures shows a slight positive deviation from additivity. Thus the equivalent conductivity, according to equation (4), will show a small negative deviation from additivity.

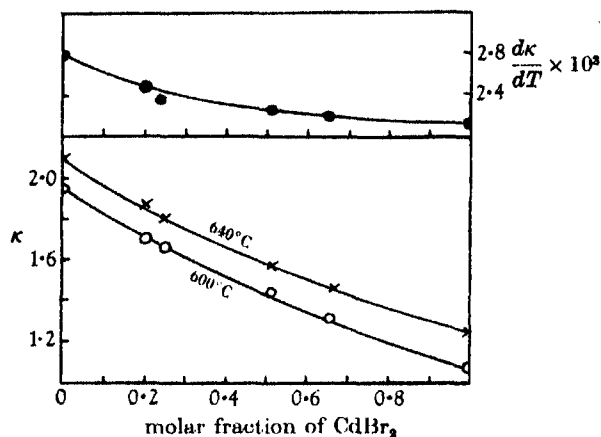


FIGURE 3. CdCl_2 - CdBr_2 . *Top*: temperature coefficient $d\kappa/dT$ against molar fraction of CdBr_2 ; *bottom*: specific conductivity (κ) against molar fraction of CdBr_2 at 600 and 640°C.

The plot of $\log \kappa$ against $1/T$ gives straight lines for all mixtures. The activation energies (C) in kcal./mol. are shown in table 2.

TABLE 2

mol. % CdBr_2	0	20.4	24.8	51.2	65.9	100
C	2.30	2.30	2.30	2.45	2.70	2.75

From the previous discussion, cadmium halides are expected to be predominantly anion conductors. A small amount of substitution of Cl^- by Br^- does not affect C much, probably because in this range conductance is due mainly to the smaller chloride ion in preference to the larger bromide ion. However, as the fraction of bromide increases further, C will gradually rise because conductance is taken over to an increasing extent by the bromide ion as the composition of the pure cadmium bromide is approached.

5. THE SYSTEM LEAD CHLORIDE-CADMIUM CHLORIDE

The phase diagram (figure 4) is of the simple eutectic type; there is no evidence of intermediate phases in the solid state (Sandonnini 1920).

At all temperatures, the specific conductivity (κ) plotted against molar fraction shows a positive deviation from additivity (figure 4).* The temperature coefficient shows a slight positive deviation.

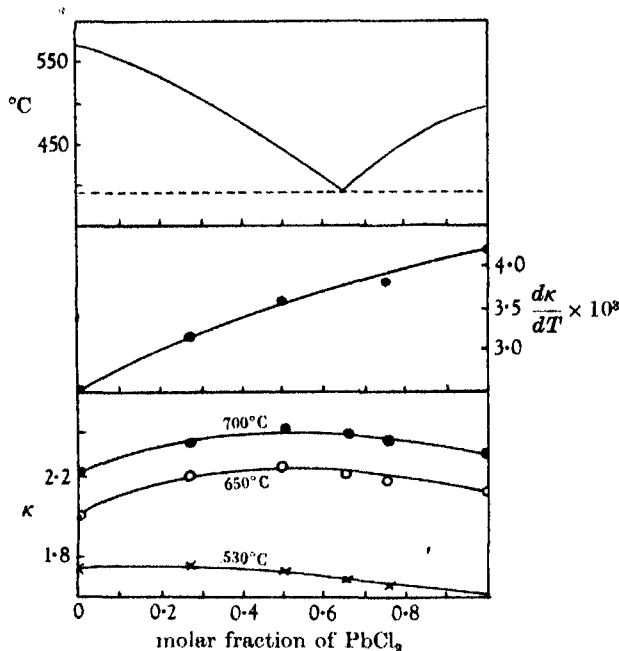


FIGURE 4. CdCl_2 - PbCl_2 . *Top*: phase diagram; *centre*: temperature coefficient $d\kappa/dT$ against molar fraction of PbCl_2 ; *bottom*: specific conductivity (κ) against molar fraction of PbCl_2 at 530, 650 and 700 $^{\circ}\text{C}$.

The activation energies (C) calculated from the plot of $\log \kappa$ against $1/T$ are shown in table 3.

The activation energy of ionic migration (C) of pure cadmium chloride is independent of temperature, the plot of $\log \kappa$ against $1/T$ being linear (figure 1). The same is found for the activation energy of viscosity (B), calculated from the viscosity determinations by Karpachev & Stromberg (1938).

Lead chloride behaves differently. The activation energy of ionic migration shows a marked increase as the melting-point is approached (cf. figure 1); the same applies to the mixtures of PbCl_2 and CdCl_2 , particularly the eutectic mixture (34.5% CdCl_2), which can be investigated at a temperature as low as 400 $^{\circ}\text{C}$.

* The equivalent conductivity, calculated from our values for κ and from preliminary determinations of the density by F. H. Dorman in this laboratory, shows also a distinct positive deviation from additivity.

The variation of C with temperature may be due to a constitutional change in molten lead chloride. The activation energy would thus contain the energy necessary to bring about that change. From the figures of table 3, it would appear that the low temperature species of lead chloride persists also in mixtures with cadmium chloride. However, the possibility of a different explanation will emerge from discussions in § 7.

TABLE 3

mol. % CdCl_2	0	24.6	34.6	49.1	72.7	100
	3.30	3.20	3.10	2.90	2.50	2.30
	(above 600°C)	(above 600°C)	(above 600°C)	(above 600°C)	(above 600°C)	(all temps.)
C	4.10	3.70	3.70	3.30	2.70	—
	(520–600°C)	(520–600°C)	(530–600°C)	(500–600°C)	(530–600°C)	—
	4.60	4.40	~ 5.0	~ 5.0	—	—
	(500–520°C)	(460–520°C)	(450–530°C)	(460–500°C)	—	—
	—	—	~ 6.0 (400–450°C)	—	—	—

The actual nature of the low temperature species of lead chloride is not known. Prasad (1933) suggested that it is the PbCl_2 molecule which is dissociated to an increasing extent on rise of temperature, but there is no conclusive evidence for a marked increase of electrovalent character on raising the temperature. Another possibility is that lead chloride may exist as autocomplexes at low temperatures which dissociate into simple ions on heating. However, transference experiments by Wirths (1937) using a radioactive indicator (ThB) gave no indication for the presence of some of the lead as part of a complex anion.

Of all systems so far examined by any investigator, the system $\text{PbCl}_2\text{-CdCl}_2$ is the only one where the conductivity of the mixtures shows a positive deviation from additivity. Moreover, the positive deviation becomes stronger on raising the temperature. No satisfactory explanation for this behaviour can be given at present.

6. THE SYSTEM CADMIUM CHLORIDE-SODIUM CHLORIDE

The phase diagram (figure 5) of this system indicates one incongruently melting compound, $\text{CdCl}_2 \cdot 2\text{NaCl}$ (Brandt 1911).

Figure 5 shows a number of isotherms of specific conductivity (κ) at various temperatures. The value of the conductivity of pure sodium chloride, which is solid at the experimental temperatures, has been extrapolated from data above the melting-point to the experimental temperatures. The isotherms show a distinct negative deviation from additivity, the amount of deviation decreasing with increasing temperature. The magnitude of the negative deviation (Δ) has a maximum at approximately the composition of the compound mentioned above. These facts are compatible with the assumption of the presence of a certain amount of $[\text{CdCl}_4]^{2-}$

ions in the mixtures, which would contribute little to the conductivity because of their large size and bivalency. The fact that the negative deviation of conductivity decreases on rise of temperature suggests that the complex dissociates into simple ions on heating.

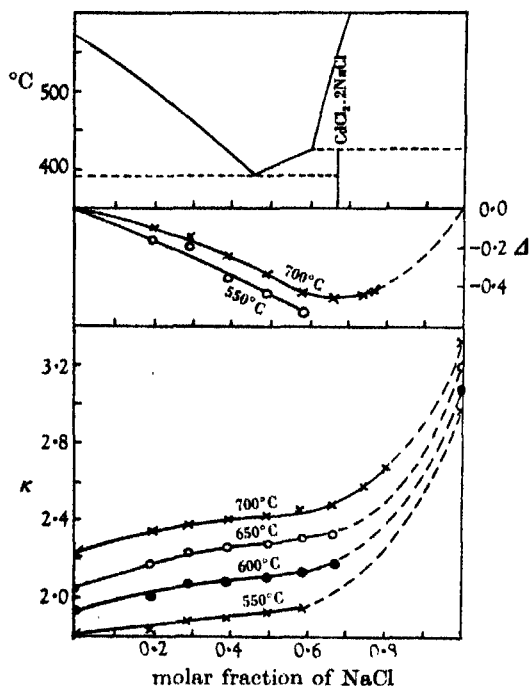


FIGURE 5. CdCl_2 -NaCl. *Top*: phase diagram; *bottom*: specific conductivity (κ) against molar fraction of NaCl at 550, 600, 650 and 700°C; *centre*: deviation Δ (negative) of specific conductivity from additivity at 550 and 700°C.

The activation energies (C) are shown in table 4 (cf. also figure 1).

TABLE 4

mol. % NaCl	0	19.1	28.8	39.1	49.0	58.4	65.9	74.0
C	2.30	2.40	2.40	2.20	2.20	2.10	2.20	2.30
	(all	(all	(above	(above	(above	(above	(above	(above
	temps.)	temps.)	650°C)	650°C)	650°C)	650°C)	650°C)	670°C)
	—	—	3.95	4.00	4.00	~5	—	~4.2
			(below	(below	(below	(below		(657—
			500°C)	500°C)	450°C)	450°C)		667°C)

In the high temperature range, there is little variation of C and $d\kappa/dT$ with composition. It will be seen in the discussions of the next two systems that the presence of dissociating complex ions gives rise to high values of C at the composition at which the fraction of such ions is at a maximum. The absence of such an effect in this system suggests that either the complex ions are dissociated to a large extent, or

sodium ions have taken over the major part of the conductance from the chloride ions at the composition corresponding to that of the complex. It is true that the activation energies have high values in the mixtures as the crystallization temperatures are approached, but significantly not only at the composition of the compound. This effect may have a different cause and will be discussed in connexion with the next two systems.

7. THE SYSTEM CADMIUM CHLORIDE-POTASSIUM CHLORIDE

This system was previously investigated by Sandonnini (1920), but only at one temperature (800° C). It is very interesting as the phase diagram (figure 6) indicates one congruently melting compound, $\text{CdCl}_2 \cdot \text{KCl}$, and one incongruently melting compound, $\text{CdCl}_2 \cdot 4\text{KCl}$, in the solid state.

We measured the specific conductivities from 720° C down to the crystallization temperatures (table 5). As the densities in this system were known due to the

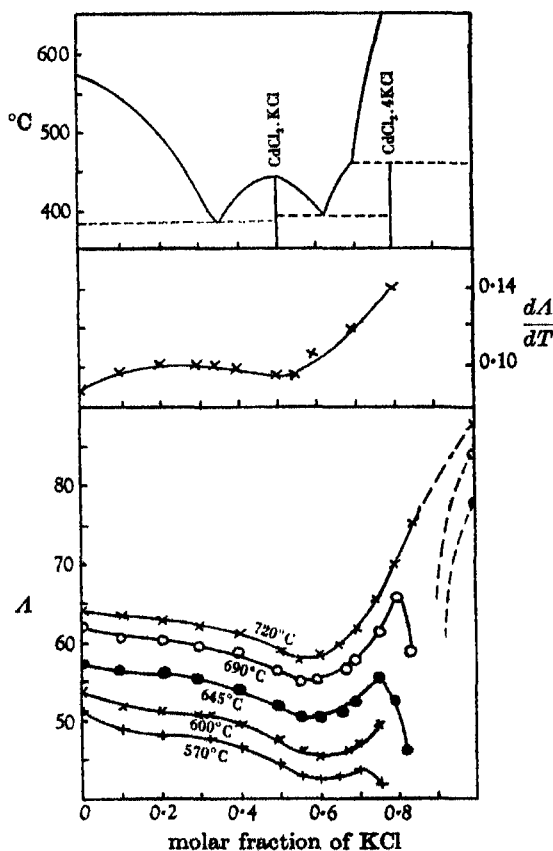


FIGURE 6. $\text{CdCl}_2\text{-KCl}$. Top: phase diagram; centre: temperature coefficient dA/dT against molar fraction of KCl; bottom: equivalent conductivity (A) against molar fraction of KCl at 570, 600, 645, 690 and 720° C.

TABLE 5. SPECIFIC CONDUCTIVITY—CdCl₂-KCl

mol. % KCl	570° C	600° C	645° C	690° C	720° C
0	1.862	1.944	2.069	2.188	2.257
9.5	1.715	1.807	1.947	2.073	2.148
20.3	1.606	1.698	1.838	1.961	2.039
29.0	1.521	1.605	1.732	1.854	1.925
32.4	1.495	1.578	1.701	1.815	1.885
39.4	1.429	1.506	1.621	1.727	1.794
49.4	1.305	1.377	1.481	1.579	1.637
54.0	1.232	1.300	1.407	1.510	1.580
59.2	1.183	1.258	1.366	1.470	1.532
66.5	1.150	1.227	1.343	1.450	1.518
69.2	1.148	1.223	1.326	1.446	1.515
74.8	—	1.260	1.380	1.496	1.568
79.6	very low	—	1.275	1.557	1.634
83.9	very low	very low	very low	1.361	1.702
100*	(1.706)	(1.770)	(1.864)	(1.954)	(1.921)

* Conductivity values extrapolated.

investigations by Lorenz and co-workers (1907, 1928), it was possible to calculate equivalent conductivities (Λ) defined as

$$\Lambda = \frac{\kappa \times E_m}{d}, \quad (6)$$

where d is the density and E_m the mean equivalent weight, the latter being defined as

$$E_m = E_1 f_1 + E_2 f_2,$$

where E_1 and E_2 are the equivalent weights and f_1 and f_2 the equivalent fractions of the components I and II (in the present case CdCl₂ and KCl).

The equivalent conductivities, plotted against the molar fractions of KCl, are shown in figure 6.† It can be seen that, on addition of KCl to CdCl₂, the equivalent conductivity decreases, even though that of potassium chloride is higher than that of cadmium chloride. This suggests the removal of Cd⁺⁺ and in particular of Cl[—]—since CdCl₂ is predominantly an anion conductor—from the melt to form an anionic complex. This complex will contribute little to the conductivity because of its large size.

For additions of KCl up to 40 mol. %, the negative deviation from additivity (Δ) is approximately proportional to the molar fraction of KCl (f_2), as shown in table 6.

This suggests that the addition of KCl to CdCl₂ will lead to the formation of a complex ion; the phase diagram would suggest [CdCl₃][—]. For each two Cl[—] thus being made unavailable for conductance, one K⁺ ion will become available. As we do not know the relative contributions of K⁺, Cl[—] and Cd⁺⁺ in these mixtures, quantitative

† The Λ -values for pure KCl were extrapolated as mentioned in § 6. It may also be mentioned that the isotherm calculated from Sandonnini's measurements of κ at 800° C (cf. Mulcahy & Heymann) is similar in shape to our isotherm at 720° C.

speculations would be of doubtful value. Moreover, it is possible that complex structures, more highly polymeric than $[\text{CdCl}_3]^-$, may exist in the molten mixtures (cf. later).

TABLE 6

f_2	720° C		700° C		600° C	
	Δ	Δ/f_2	Δ	Δ/f_2	Δ	Δ/f_2
0.10	3.2	32.0	3.0	30.0	3.3	33.0
0.20	6.3	31.5	6.2	31.0	5.7	28.5
0.30	9.4	31.3	9.4	31.3	9.5	31.7
0.40	13.0	32.2	13.0	32.2	12.0	30.0

If $[\text{CdCl}_3]^-$ were the only complex species present in the melt, a minimum in the isotherm (or rather a maximum of the negative deviation) would be expected near a mole fraction of 0.5. However, the minimum occurs at $f_2 > 0.50$. This is probably due to the presence of another complex, containing more chlorine, in the melt. The phase diagram, indicating an incongruently melting compound, $\text{CdCl}_2 \cdot 4\text{KCl}$, suggests a complex of composition $[\text{CdCl}_6]^{4-}$. Evidence for the existence of a complex of this composition in the molten mixtures is seen in the fact that the isotherms tend towards a deep minimum at $f_2 = 0.8$ at a temperature of 690° and below. At higher temperatures, however, no minimum is found at this composition. Thus the complex $[\text{CdCl}_6]^{4-}$ may be assumed to be highly exothermic and apparently dissociates completely some 30 to 40° above the melting point of the compound $\text{CdCl}_2 \cdot 4\text{KCl}$.

The relative instability of the complex $[\text{CdCl}_6]^{4-}$ is also indicated by the fact that the temperature coefficient of the equivalent conductivity ($d\Lambda/dT$) increases towards the composition of the compound $\text{CdCl}_2 \cdot 4\text{KCl}$ (figure 6). The high value of $d\Lambda/dT$ is probably due to the fact that $[\text{CdCl}_6]^{4-}$ of low mobility dissociates on heating with the formation of Cd^{++} and Cl^- , which have a greater mobility.

These results favour the conclusion that, even if an intermediate phase of a binary salt system has an incongruent melting-point, the complex ion corresponding to that compound may persist in the molten state in a limited range of temperature.

The activation energies C are shown in table 7.

The activation energy C^* in the high temperature range has a maximum around the composition of the compound $\text{CdCl}_2 \cdot 4\text{KCl}$, suggesting that C includes part of the heat of dissociation of the complex, which may be represented as



At temperatures approaching the crystallization points, C increases sharply. In the range between 70 and 80 mol. %, it goes up to 8 to 9 kcal. at temperatures of 10 to 20° C above the crystallization temperature. High values for the activation energy just above the crystallization temperature are an expression of the fact that the curves of κ against temperature show a strong downward slope in that region

* The activation energy C' was also calculated (from equation (3a)), although over a smaller range of temperature in view of the scarcity of density data in some temperature ranges. The variation of C' with composition is very similar to that of C . This applies also to the system $\text{PbCl}_2\text{-KCl}$ (cf. § 8).

(figure 7, cf. also table 5). This phenomenon is not confined only to the composition corresponding to that of an incongruently melting compound but occurs at other compositions also, as mentioned in the systems $\text{CdCl}_2\text{-NaCl}$ and $\text{PbCl}_2\text{-CdCl}_2$. The highest values of C just above the crystallization temperature are not far below those

TABLE 7

mol. % KCl	9.5	20.3	29.0	32.4	39.4	49.4	
C	2.67 (all temps.)	2.70 (all temps.)	2.70 (all temps.)	2.67 (all temps.)	2.65 (all temps.)	2.55 (above 620°C)	
	—	—	—	—	—	2.95 (570–620°C)	
mol. % KCl	54.0	59.2	66.5	69.2	74.8	79.6	83.9
C	2.67 (above 620°C)	2.85 (570–720°C)	3.22 (570–720°C)	3.22 (570–720°C)	3.22 (620–720°C)	3.25 (680–720°C)	2.30 (704–725°C)
	2.90 (570–620°C)	—	—	—	~ 5 (600–620°C)	3.45 (660–680°C)	~ 9 (690–704°C)
	3.22 (470–560°C)	—	—	—	~ 9 (590–600°C)	~ 8 (645–660°C)	—
	~ 7 (430–460°C)	—	—	—	—	—	—

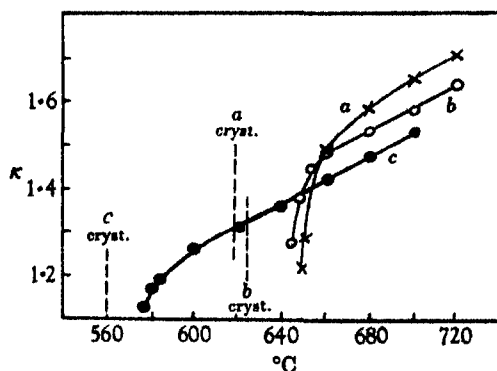


FIGURE 7. Plot of specific conductivity (κ) against temperature. (a) 20.5 mol. % PbCl_2 + 79.5 mol. % KCl ; (b) 20.4 mol. % CdCl_2 + 79.6 mol. % KCl ; (c) 25.2 mol. % CdCl_2 + 74.8 mol. % KCl . The vertical broken lines indicate the crystallization temperature of these mixtures.

frequently found for activation energies of ionic migration in crystalline solids. This may be interpreted as being due to the establishment of an increasing degree of order as the crystallization temperature is approached. Certain X-ray and specific heat effects with ordinary liquids just above their melting-points may be explained

in similar terms (cf. J. Frenkel and other contributors in the Faraday Society Discussion on 'The Liquid State' (1937) and Randall (1934)).

In the case of the compound $\text{CdCl}_2 \cdot \text{KCl}$, a certain degree of order in the melt may be explained in more specifically chemical terms. According to the X-ray studies of Brasseur & Pauling (1938), each cadmium atom in solid salts of the type KCdCl_3 is surrounded by six chlorine atoms in the form of a regular octahedron. Each octahedron shares two opposite edges with similar octahedra to form chains of $[\text{CdCl}_3]^-$ complexes, which occur in the crystal in the form of infinite polymers. At the melting-point, these chains will break up, but the disordered lattice of the melt may still contain short chains, which should, however, be regarded as statistical rather than actual units. On further absorption of heat, disorder will increase and the chains will break up further. Such a model leads us to expect the low values of conductivity and high activation energies just above the melting-point which have been found experimentally.

Future theoretical investigations may treat these phenomena in terms of the absolute theory of rate processes as applied to viscosity and similar phenomena (Eyring 1936; Glasstone, Laidler & Eyring 1941). The increase of the (total) energy of activation on lowering the temperature may be due mainly to an increase of the entropy of activation. As the unit of ionic migration is the ion, the entropy of activation will be greater for the somewhat ordered melt near the melting-point than for the more disordered melt at a higher temperature.

8. THE SYSTEM LEAD CHLORIDE-POTASSIUM CHLORIDE

The phase diagram of this system (figure 8) indicates a congruently melting compound, $2\text{PbCl}_2 \cdot \text{KCl}$, and two incongruently melting compounds, $\text{PbCl}_2 \cdot 2\text{KCl}$ and $\text{PbCl}_2 \cdot 4\text{KCl}$ (Lorenz & Ruckstuhl 1906). The presence of anionic complexes in the molten state was proved very convincingly by transference experiments by Wirths (1937), who used a radioactive lead isotope (Th B).

In the calculation of the equivalent conductivity (Λ) from the specific conductivity (κ), shown in table 8, the density data by Lorenz, Frei & Jabs (1907) were used.*

The isotherms of equivalent conductivity (figure 8) can be understood on the assumption of the presence of part of the lead in an anionic complex. On addition of potassium chloride to lead chloride, the equivalent conductivity falls, although that of KCl is higher than that of PbCl_2 . Addition of KCl obviously removes Pb^{++} and Cl^- from the melt with the formation of anionic complexes, which contribute but little to the conductance. This behaviour is similar to that in the system $\text{CdCl}_2 \cdot \text{KCl}$.

Up to 45 to 50 mol. %, the negative deviation from additivity is approximately proportional to the molar fraction of KCl . Moreover, a minimum is found in the isotherms at $f_2 = 0.55$ at high temperatures which shifts towards $f_2 = 0.60$ at lower temperatures. The proximity of the minimum to $f_2 = 0.5$ would suggest that the

* Regarding the Λ -value of pure KCl , cf. § 7.

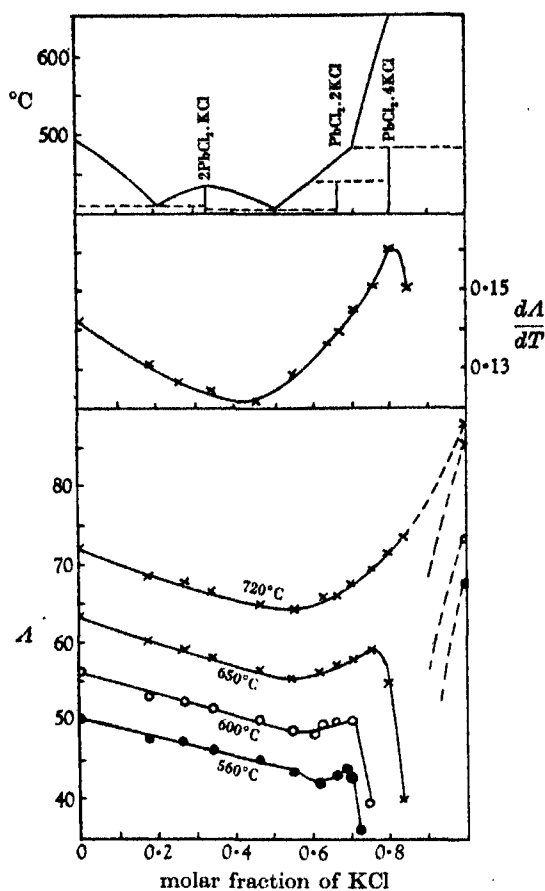


FIGURE 8. PbCl_2 -KCl. Top: phase diagram; centre: temperature coefficient $d\Lambda/dt$ against molar fraction of KCl; bottom: equivalent conductivity (Λ) against molar fraction of KCl at 560, 600, 650 and 720° C.

TABLE 8. SPECIFIC CONDUCTIVITY— PbCl_2 -KCl

molar fraction of KCl	560° C	600° C	650° C	700° C	720° C
0	1.75	1.92	2.13	2.34	2.39
17.8	1.58	1.72	1.90	2.07	2.13
26.9	1.50	1.64	1.80	1.96	2.01
34.0	1.43	1.56	1.72	1.87	1.92
36.3	1.41	1.53	1.69	1.84	1.88
46.2	1.32	1.43	1.58	1.73	1.77
54.8	1.22	1.35	1.50	1.64	1.69
62.1	1.18	1.31	1.46	1.62	1.66
66.2	1.15	1.30	1.46	1.60	1.65
67.3	1.18	1.30	1.45	1.60	1.65
70.3	1.17	1.29	1.45	1.61	1.67
75.1	~ 0.6	1.00	1.46	1.61	1.67
79.5	—	—	1.22	1.64	1.69
83.5	—	—	~ 1.0	1.67	1.73
100	(1.69)	(1.77)	(1.88)	(1.98)	(2.02)

stable complex ion in the melt is $[\text{PbCl}_3]^-$ or a polymeric complex structure of similar composition. The minimum near $f_2 = 0.5$ is somewhat surprising because the only stable solid compound indicated by the phase diagram corresponds to $f_2 = 0.33$ ($2\text{PbCl}_2 \cdot \text{KCl}$), and it is doubtful whether there is an anhydrous solid compound of composition $\text{PbCl}_2 \cdot \text{KCl}$. Moreover, transference experiments by Tubandt & Reinhold (1923) showed that the transport number of chlorine in the solid $2\text{PbCl}_2 \cdot \text{KCl}$ was equal to one, suggesting that crystals of this compound do not contain complex ions (cf. also the crystal structure determinations by X-rays by Mehmehl & Nespital (1934)). In the light of these results, it would appear—though conditions of stability of a complex in the liquid state at high temperature may be different from those of the solid—that the complexes in the molten state may have to be regarded as statistical rather than discrete units, and that the bonding between the metal atom and the chlorine atoms of the 'complex' may be largely electrovalent in nature.

Unstable complexes, containing more chlorine than $[\text{PbCl}_3]^-$, are likely to exist in molten mixtures containing a high proportion of KCl . This is indicated by the very sharp second minimum towards $f_2 = 0.8$ which is found at a temperature of 650° and below. Above that temperature, however, the second minimum does not occur. This suggests an unstable complex of composition $[\text{PbCl}_6]^{4-}$, corresponding to the compound $\text{PbCl}_2 \cdot 4\text{KCl}$, which dissociates on rise of temperature.

Figure 8 also shows the variation of dA/dT with composition, dA/dT having a minimum between 30 and 50 mol. % and a distinct maximum at 80 mol. %. The activation energies (C), calculated* as usual from the plot of $\log \kappa$ against $1/T$, are shown in table 9.

TABLE 9

mol. % KCl	0	17.8	26.9	34.0	46.2	54.8
C	3.30 (above 600°C)	3.00 (above 640°C)	3.10 (560– 720°C)	3.10 (560– 720°C)	3.10 (560– 720°C)	3.45 (640– 720°C)
	4.10 (520– 600°C)	3.45 (560– 640°C)	—	—	—	3.80 (560– 620°C)
	4.60 (500– 520°C)	—	—	—	—	—
mol. % KCl	62.1	66.2	70.3	75.1	79.5	83.5
C	3.45 (640– 720°C)	3.45 (640– 720°C)	3.82 (640– 720°C)	3.80 (640– 720°C)	3.80 (680– 720°C)	3.22 (710– 730°C)
	3.63 (560– 640°C)	3.86 (560– 640°C)	3.96 (580– 640°C)	~ 7.5 (620– 640°C)	4.70 (660– 680°C)	—
	—	—	—	—	~ 12 (650– 660°C)	—

* Regarding the activation energies C' , cf. § 7.

At high temperatures (around 700°), C has a maximum near the compositions where the phase diagram indicates incongruent compounds, suggesting that C includes part of the energy of dissociation of the corresponding complexes.*

At low temperatures when the crystallization points are approached, very high values for C are found, particularly in the region between 70 and 80 mol. % of KCl. This behaviour is similar to that found in the system $\text{CdCl}_2\text{-KCl}$ and may similarly be explained as being due to an increasing degree of order in the melt as the crystallization temperature is approached. This is also suggested by the curves of κ against temperature (figure 7, cf. also table 8), which show a distinct increase of slope in the liquid state near the melting-point.

9. RELATIONS BETWEEN THE CONSTITUTION OF MIXTURES OF MOLTEN SALTS AND PROPERTIES RELATED TO THEIR CONDUCTIVITY

In no system so far investigated is the conductivity of the mixtures strictly additive. Even in systems in which there is no evidence of compound formation, the isotherms of equivalent conductivity show moderate negative deviations from additivity,† viz. NaCl-KCl , NaCl-CaCl_2 , NaCl-BaCl_2 , NaCl-PbCl_2 (Barzakowski 1940); LiCl-KCl , $\text{KNO}_3\text{-NaNO}_3$ (Klochko 1940); $\text{PbCl}_2\text{-PbBr}_2$, AgCl-PbCl_2 (cf. Mulcahy & Heymann 1943); $\text{CdCl}_2\text{-CdBr}_2$ (§ 4). So far, no satisfactory explanation for these deviations from additivity in simple systems has been given (Mulcahy & Heymann 1943).

In systems which give evidence of compound formation in the solid state, the conductivity of the molten mixtures always shows strong negative deviations from additivity. In systems in which the phase diagram gives evidence of only one compound, the maximum deviation occurs near the composition of that compound, viz. $\text{CdCl}_2\text{-2NaCl}$ (§ 6) and $\text{CaCl}_2\text{-KCl}$ (cf. Mulcahy & Heymann). In systems in which the phase diagram indicates several compounds, viz. $\text{CdCl}_2\text{-KCl}$ (§ 7) and $\text{PbCl}_2\text{-KCl}$ (§ 8), more than one minimum is sometimes found in the conductivity isotherms at low temperatures. Those corresponding to incongruently melting compounds, however, disappear as the temperature is raised ($\text{CdCl}_2\text{-KCl}$, $\text{PbCl}_2\text{-KCl}$). The position of the minimum may not correspond to the composition of the respective compound in the solid state, suggesting that the conditions of stability in the liquid state may be different from those in the solid state.

* It can be shown easily that, at constant temperature, the activation energies are proportional to the relative temperature coefficients (viz. $\frac{1}{\kappa} \frac{d\kappa}{dT}$), whereas the absolute temperature coefficients (viz. $\frac{d\kappa}{dT}$) are proportional to $C\kappa$ and CA respectively.

In the systems $\text{CdCl}_2\text{-KCl}$ and $\text{PbCl}_2\text{-KCl}$, dA/dT has a minimum near the composition of the stable compound, although C and C' show no such minimum. This is due to the fact that A has a minimum at the composition of the stable compound, and that dA/dT is proportional to CA .

† The only exception so far is the system $\text{CdCl}_2\text{-PbCl}_2$, where the conductivity shows a positive deviation from additivity (cf. § 5).

In the molten systems $\text{CdCl}_2\text{-KCl}$ and $\text{PbCl}_2\text{-KCl}$, the presence of unstable complexes which undergo dissociation on heating is also indicated by high values of the activation energy (C) near compositions corresponding to compounds which may yield such complex ions, viz. $[\text{CdCl}_6]^{4-}$ and $[\text{PbCl}_6]^{4-}$. In these cases, C contains part of the energy of dissociation of the complex in question. At a composition corresponding to a stable complex, viz. $[\text{CdCl}_3]^-$ and perhaps $[\text{PbCl}_3]^-$, however, no such maximum of C is found. Apparently these complexes do not dissociate appreciably on heating.

In the system KCl-MgCl_2 , negative deviations but no distinct minima have been recorded (Klochko 1940; Shcherbakov & Markov 1939), although the phase diagram indicates two compounds with congruent melting-points. However, the experimental temperatures were about 200° above the melting-points of the compounds, and the corresponding complexes may be largely dissociated at these temperatures. This conclusion receives support from the fact that the relative temperature coefficient of conductivity $\left(\frac{1}{\kappa} \frac{d\kappa}{dT}\right)$ which is proportional to C has a distinct maximum near the composition of the compound $2\text{KCl} \cdot \text{MgCl}_2$, suggesting that the corresponding complex dissociates appreciably on rise of temperature.

We wish to thank M. F. R. Mulcahy for interesting discussions.

REFERENCES

- Andrade 1934 *Phil. Mag.* (7), 17, 698.
Aten 1910 *Z. phys. Chem.* 73, 580.
Barrer 1943 *Trans. Faraday Soc.* 39, 48.
Barzakowski 1940 *Bull. Acad. Sci. U.R.S.S., Class Chim.* no. 5, p. 825.
Biltz & Klemm 1926 *Z. anorg. Chem.* 152, 267.
Brandt 1911 *N. Jb. Min. Geol. Paläont.* 32, 630.
Brasseur & Pauling 1938 *J. Amer. Chem. Soc.* 60, 2886.
Dunn 1926 *Trans. Faraday Soc.* 22, 401.
Espe & Knoll 1936 *Werkstoffkunde der Hochvakuumtechnik*. Berlin: J. Springer.
Eyring 1936 *J. Chem. Phys.* 4, 283.
Frenkel 1937a *Trans. Faraday Soc.* 33, 58.
Frenkel 1937b *Acta Physicochim. U.R.S.S.* 6, 341.
Glasstone, Laidler & Eyring 1941 *The theory of rate processes*, pp. 477-552. New York: McGraw-Hill Co.
Karpachev & Stromberg 1934 *J. Phys. Chem. U.S.S.R.* 5, 793.
Karpachev & Stromberg 1935 *J. Gen. Chem. U.S.S.R.* 5, 5.
Karpachev & Stromberg 1938 *J. Phys. Chem. U.S.S.R.* 11, 852.
Klochko 1940 *Bull. Acad. Sci. U.R.S.S., Class Chim.* no. 5, p. 640.
Lorenz & Adler 1928 *Z. anorg. Chem.* 172, 372.
Lorenz, Frei & Jabs 1907 *Z. phys. Chem.* 61, 468.
Lorenz & Kalmus 1907 *Z. phys. Chem.* 59, 17, 244.
Lorenz & Ruckstuhl 1906 *Z. anorg. Chem.* 51, 70.
Mehmel & Nespital 1934 *Z. Kristallogr.* 88, 345.
Mott & Gurney 1940 *Electronic processes in ionic crystals*. Oxford University Press.
Mulcahy & Heymann 1943 *J. Phys. Chem.* 47, 485.
Nacken 1907 *Zbl. Min. Geol. Paläont.*, 262.

- Pauling 1940 *The nature of the chemical bond*, p. 40. New York: Cornell University Press.
Prasad 1933 *Phil. Mag.* **16**, 263.
Randall 1934 *The diffraction of X-rays and electrons by amorphous solids, liquids and gases*.
London: Chapman and Hall.
Sandonnini 1920 *Gaz. Chim. Ital.* **50**, 289.
Sheherbakov & Markov 1939 *J. Phys. Chem. U.S.S.R.* **13**, 353, 621.
Tubandt & Reinhold 1923 *Z. Elektrochem.* **29**, 213.
Tubandt & Reinhold 1925 *Z. Elektrochem.* **31**, 84.
Tubandt, Reinhold & Liebold 1931 *Z. anorg. Chem.* **197**, 225.
Ward 1937 *Trans. Faraday Soc.* **33**, 88.
Wirths 1937 *Z. Elektrochem.* **43**, 486.
-

An electron diffraction study of the structure of thin films of normal paraffins

By K. G. BRUMMAGE

(Communicated by Sir Frank Smith, F.R.S.—Received
14 March 1946.—Read 23 May 1946)

[Plates 13 and 14]

Thin films of the normal paraffins tetracosane ($C_{24}H_{50}$), triacontane ($C_{30}H_{62}$) and tetratriacontane ($C_{34}H_{70}$) have been examined by electron diffraction when deposited on stainless steel, copper or collodion. Both the reflexion and the transmission techniques have been employed. The results of rubbing the films have been ascertained, and the orientations observed have been explained in terms of the crystal structure of the normal paraffins.

Films of the three paraffins have been heated upon stainless steel and upon copper surfaces, and curves derived relating the temperature at which the ordered film structure breaks down, to the film thickness. These results suggest that this temperature corresponds in the case of a monolayer to a transition of the film structure from a condensed to an expanded type.

INTRODUCTION

When two lubricated surfaces pass over each other under sufficiently heavy load, the lubricant film between them becomes extremely thin and hydrodynamic conditions of lubrication no longer apply. Lubrication is then effected by the adsorbed films on the rubbing surfaces, and the behaviour of such films under shear and during heating is of considerable interest in the fundamental study of boundary lubrication.

Such films may conveniently be studied by electron diffraction methods. Owing to the complexity of commercial lubricating oils, it is much simpler to examine, in the first place, pure materials, typical of those to be found in lubricants. The present paper deals with the simplest constituents, the normal paraffins; it is hoped subsequently to deal with other types of pure compounds and so to obtain a general picture of the behaviour of lubricating oil constituents.

It has been found that at room temperatures, films of normal paraffins such as *n*-tetracosane, *n*-triacontane and *n*-tetratriacontane are invariably crystalline, both before and after subjection to sliding shear. The molecular positions observed relative to the surface are completely determined by the crystal structure. When such films formed on metallic surfaces are heated, they lose their orderly arrangement at some temperature which depends upon molecular composition and film thickness. This disorientation temperature is lowest for the monolayer and tends towards the bulk melting-point with increasing film thickness. In films only one molecule thick, this disorientation may be considered as a transition of the adsorbed paraffin film from a condensed to an expanded state.

APPARATUS

The electron diffraction camera employed was of the Finch type, and has been described by Finch & Wilman (1937). The main addition to this equipment was a reflexion fitting which could be heated electrically in the camera whilst the diffraction pattern from the film upon the surface was being examined. This was provided with interchangeable stainless steel and copper specimen surfaces; the temperature at the surface being examined was measured by means of an iron-eureka thermocouple. It could be heated to at least 250° C without damage. The whole assembly fitted into a standard camera port with the usual ground and greased joint, the electrical leads being brought out of the camera through bushes in the specimen carrier.

PREPARATION OF REFLEXION SPECIMENS

Films of the normal paraffins tetracosane ($C_{24}H_{50}$), triacontane ($C_{30}H_{62}$) and tetratriacontane ($C_{34}H_{70}$) have been examined on metal surfaces and on thin films of collodion, using the transmission technique in the latter case. The metal surfaces used were stainless steel and copper, polished by conventional metallurgical methods, and electrolytically polished stainless steel, all in the form of 1 in. disks. The electrolytic polish was obtained using a bath of two parts of acetic anhydride and one of perchloric acid, as described by Jacquet & Rocquet (1939).

Electron diffraction patterns from both the stainless steel surfaces showed no signs of preferential scattering, despite the crystalline nature of the electrolytically polished surface. This lack of pattern was probably due to the flatness of the surfaces.

Patterns from the copper disclosed the presence of air-formed films of cuprous oxide or sulphide. It is well recognized that, with few possible exceptions, no metallic surface exposed to air is composed simply of the bare metal. Any oil film must be present on top of a surface film of oxide or sulphide, and this applies as much to the surfaces used in engineering practice, as to those employed in the present experiments. It must be remembered, too, that even the finest mirror finish is quite rough on a molecular scale.

Films for reflexion purposes were formed by placing a known quantity of a solution in isohexane of one of the paraffins on one of the reflexion disks, and adding sufficient pure solvent to spread the solution over the whole surface. The specimen was then rotated at about 70 r.p.m. about a vertical axis through the centre of the disk, while the solvent evaporated. To retard the evaporation, the rotating specimen was enclosed in a 5 in. cube wooden box, the atmosphere in which was kept nearly saturated with solvent vapour. In this way, it was believed that an even continuous film of paraffin would be deposited; subsequent observations have confirmed this.

The diffraction patterns from different film thicknesses were substantially the same, and from them the area of surface occupied by a single molecule in a monomolecular layer was found to be 18.5 \AA^2 . Having deduced this, the quantity of solution deposited on the surface could be regulated to give any chosen number of molecular layers.

THE STRUCTURE OF UNRUBBED FILMS ON METALS

Figure 1, plate 13, is a good example of the patterns obtained. The same pattern was obtained in all azimuthal directions, so that the paraffin was orientated in such a way that its structure was symmetrical about an axis normal to the surface. Consideration of the spacing of the horizontal zones upon which the diffractions were grouped led to the conclusion that they were successive orders of diffractions by the alternate carbon-carbon spacing of 2.54 \AA in the zigzag n -paraffin molecule. The planes of the molecules were normal to the surface, and in all azimuths relative to the incident beam.

It can be deduced that for a film composed of independent molecules arranged in this manner, the intensity distribution along a zone or layer line is proportional to $(1 \pm J_0(Bx \tan \phi))$, where x is the distance along a layer line, ϕ is the angle between the lines joining the end carbon atom of a molecule to the two neighbouring carbon atoms in the chain, and $B = \pi a / \lambda L$, a being the distance between alternate carbon atoms in the chain, λ the wave-length of the electron beam and L the camera length from specimen to photographic plate. The positive sign applies to even order zones and the negative to odd orders. This expression was given by Germer & Storks (1938) and applied by them to barium stearate patterns.

The intensity distributions in figure 1 are in good agreement with the theoretical ones, but in addition to the diffuse zone pattern, there is on figure 1 a strong pattern of vertical streaks, indicating the existence of crystal structure in the film. All the observed streaks are explicable in terms of the n -paraffin structure given by Muller (1928), in which the unit cell is orthorhombic, with $a = 7.45 \text{ \AA}$, $b = 4.97 \text{ \AA}$ and c about twice the length of a molecule, the molecular axes being parallel to the c -axis of the unit cell. Since the diffuse pattern indicates that the molecular planes are normal to the surface and in random azimuth, the crystals into which the molecules are grouped must be in random (001) orientation on the surface.

Crystal diffractions can occur only where the diffuse pattern due to the independent molecules is strong, and a study of the reciprocal lattice of the n -paraffin crystals leads to the conclusion that the streaks must occur on the diffuse horizontal zones at distances, s , from their centres, where s is given by

$$s = \lambda L \sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)}.$$

In this equation, h and k are Laue indices relative to the a - and b -axes respectively. The closeness of the agreement is shown in table 1, where the mean values of $s/\lambda L$ from about 100 patterns are compared with the calculated values of $\sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)}$, using $a = 7.45 \text{ \AA}$ and $b = 4.97 \text{ \AA}$.

TABLE 1. OBSERVED DIFFRACTIONS IN THE NORMAL PARAFFIN REFLEXION PATTERN

layer line order	Laue indices h, k	$s/\lambda L$	$\sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)}$
2	00	0	0
1	01	0.20	0.20
1 and 2	11	0.24	0.24
1 and 2	20	0.27	0.27
1	21	0.34	0.33
2	02	0.40	0.40
1	12	0.43	0.42
1 and 2	31	0.46	0.45
1	22	0.50	0.48
1	32	0.57	0.57
1	23	0.66	0.66

The (001) orientation of the n -paraffin crystals having been thus established, it follows immediately from a knowledge of the positions of the four molecules in the unit cell that the area occupied by one molecule on the surface is $\frac{1}{2}(7.45 \times 4.97) \text{ \AA}^2$, or about 18.5 \AA^2 , as stated on p. 416.

In figure 1, the diffraction patterns due to the crystals and to the ungrouped molecules exist together. Two explanations are possible. Either the films consisted of a mixture of crystals and isolated molecules, or they were completely crystalline with the molecules executing thermal motions within the lattice. In view of the pronounced crystalline structure of the paraffins and of the weakness of the Van der Waals bonding between neighbouring molecules, the latter explanation is much the more probable. The production of diffuse patterns by molecules in thermal motion has been discussed by Charlesby, Finch & Wilman (1939) in the case of anthracene. The conditions in n -paraffin crystals are very similar except that complete rigidity of the molecules cannot be assumed.

The elongation of the crystal diffraction streaks is determined mainly by the length of crystal grating normal to the surface available to diffract the electrons. It corresponded in this case to about 35 \AA , the depth to which the electrons were

able to penetrate the films. The upper surfaces of the films must have been very flat on a molecular scale, otherwise projections of paraffin would have allowed access to gratings much longer than 35 Å.

The width of diffraction streaks along the layer lines is a measure of the crystal size in directions parallel to the surface. This varied with the thickness of the layer and with rubbing treatment, and is subsequently referred to simply as 'crystal size'.

The differences between the diffraction patterns from the three normal paraffins studied were small, except that for a given number of molecular layers deposited without rubbing on the same surface, the crystal size was larger with larger molecules.

For a given paraffin, crystal size increased with film thickness. With triacontane and tetratriacontane, the sharpness of the streaks was so good for a four molecular layer that little improvement was to be observed in thicker films. If a quantity equivalent to 50 layers or more was deposited, the (001) orientation broke down, and the sharpness of the pattern decreased. Rings were also observed characteristic of a film structure of randomly disposed *n*-paraffin crystals.

The influence of the substrate metal was found to be small. No difference could be detected between the patterns from films on conventionally polished stainless steel and on copper, although patterns from films on electrolytically polished stainless steel were slightly more diffuse, indicating a somewhat smaller crystal size.

In all film thicknesses, the orientation of the molecular axes normal to the plane of the surface was excellent. This is surprising when it is remembered that the surface was rough in terms of molecular dimensions. It leads to the conclusion that, whilst much of the film was in direct contact with the solid substrate, many of the molecules were supported in the crystal lattice in such a way that they bridged over the hollows in the surface.

THE STRUCTURE OF RUBBED FILMS ON METALS

Films of various thicknesses on metals were subjected to shear by placing the surfaces face downwards on clean no. 5 Whatman filter paper and moving them gently over the paper for a total distance of 45 cm. in one direction. The pressure applied had no appreciable effect upon the patterns subsequently observed. This method of rubbing was used by Finch (1938), in the study of stearic acid films.

Rubbed films were examined in the diffraction camera with the beam both along and at right angles to the rubbing direction. The principal structural changes observed were of two kinds:

- (1) Change of crystal size and perfection of orientation of the vertically orientated molecules.

- (2) Change of crystal orientation.

A small but appreciable increase in crystal size occurred in one- and two-molecular layer films. Occasionally, also, patterns similar to figure 2, plate 13,

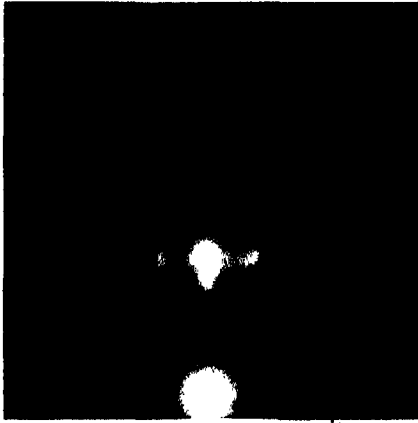


FIGURE 1

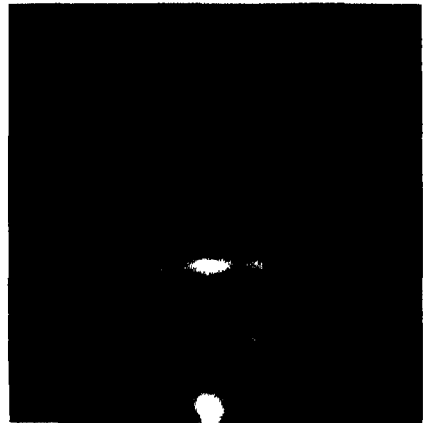


FIGURE 2

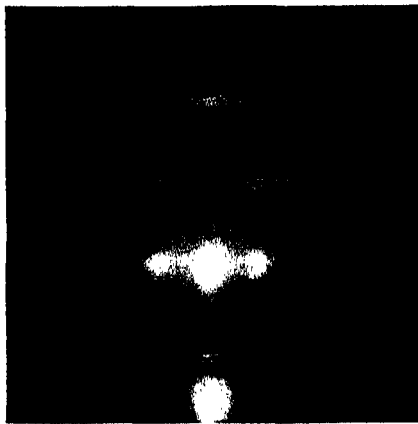


FIGURE 3

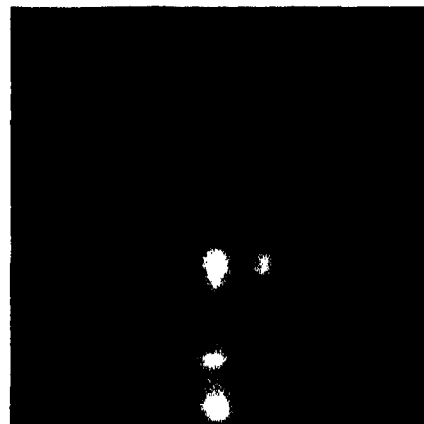


FIGURE 4

FIGURES 1-4. Reflexion electron diffraction patterns from normal paraffin films

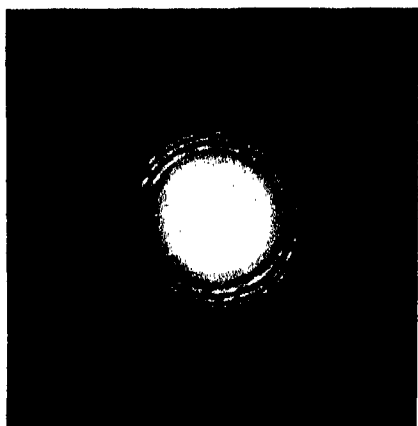


FIGURE 5

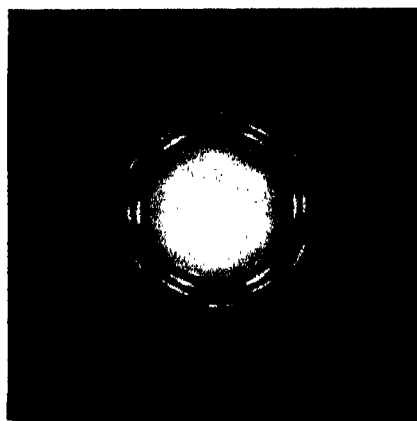


FIGURE 6

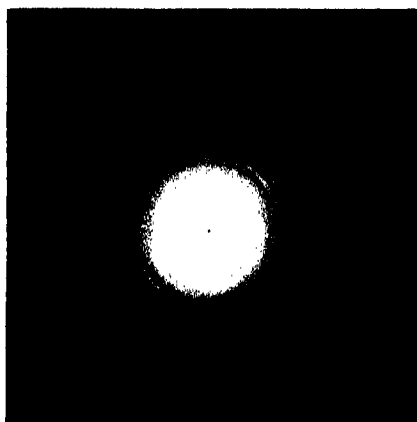


FIGURE 7

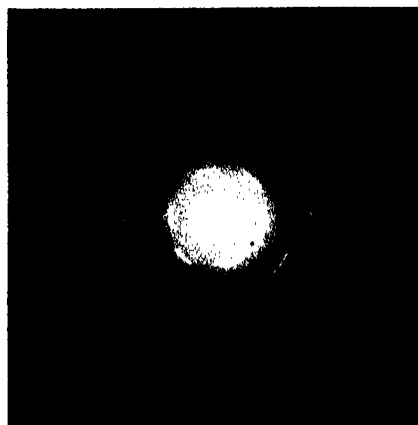


FIGURE 8

FIGURES 5-8. Transmission electron diffraction patterns from normal paraffin films.

were obtained, indicating that the layer structure had broken down to give crystals in approximate (001) orientation, but not so accurately alined as in films giving the pattern of figure 1.

In thicker rubbed films, an additional pattern of arcs was often observed superimposed on the usual layer line pattern. Figures 3 and 4, plate 13, show well-developed examples of the patterns obtained with the beam parallel and at right angles, respectively, to the direction of rubbing. In figure 4, the direction of motion of the rubbing paper across the surface was from right to left. The additional arcs in these two figures are due to *n*-paraffin crystals in which the *c*-axis is alined in the direction of rubbing, and making an angle of about 3° with the surface. This is shown by the grouping of the arcs in figure 4 upon zones spaced a distance apart corresponding to diffraction by the alternate carbon-carbon spacing of the molecules, and making an angle of 3° with the normal to the shadow edge. The 3° angle of tilt of the *c*-axis, that is of the molecular axes, always subtended the direction from which the rubbing surface approached.

The radii of the arcs on the roughly vertical zone through the undeflected spot in figure 4 show that their Laue indices are 110, 200, 220 and 400, so that the crystals had either their (110) or their (100) planes parallel to the surface, with the exception of the 3° tilt. The positions and radii of the arcs of figure 3 confirm the deductions from figure 4.

This type of crystal orientation in the rubbed films, that is mixed (110) and (100) orientations with the molecular axes alined in the rubbing direction and making and angle of 3° with the surface, was observed in rubbed films of triacontane and tetratriacontane when deposited on conventionally polished copper or stainless steel, but not on electrolytically polished stainless steel. Tetracosane did not give evidence of these orientations on any of the surfaces employed.

THE STRUCTURE OF FILMS ON COLLODION

Thin films of the three paraffins studied were examined by the transmission technique after deposition upon collodion films mounted on nickel gauze. The collodion contributed a pattern of diffuse haloes which could not be confused with the paraffin patterns. The films were formed on the collodion either by direct application of isohexane solution, or by forming the paraffin film on hot or cold water and removing it on collodion after evaporation of the solvent.

A wide variety of transmission patterns was obtained, and typical examples are shown in figures 5-8, plate 14; in all these the incident beam was normal to the specimen film. Generally, different types of patterns were yielded by different parts of the same specimen.

Photographs similar to figures 5 and 7 were usually obtained from specimens deposited direct from solution on the collodion, and from films spread on cold water. They are both due to films consisting of crystals with their *c*-axes normal to the plane of the collodion, but otherwise unorientated. In figure 7 the axes of the

molecules are all tilted slightly in one direction, due, in all probability, to sagging of the collodion between the meshes of the nickel gauze support. The (001) orientation was verified in such cases by tilting the film through 45° about an axis normal to the incident beam; the resultant pattern always consisted of arcs on a set of zones parallel within a few degrees to the axis of rotation. It follows that the molecular axes were all parallel to each other and inclined at only a small angle to the beam when this was incident normally on the film.

Since the crystals in figures 5 and 7 are in random (001) orientation, the rings must have Laue indices of the form $hk0$, and their radii are, therefore, given by

$$r = \lambda L \sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)},$$

where the symbols have their usual meanings. The agreement is, in fact, excellent as table 2 shows.

TABLE 2. RING MEASUREMENTS, RANDOM (001) TRANSMISSION PATTERN

Laue indices	$r/\lambda L$	$\sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)}$	Laue indices	$r/\lambda L$	$\sqrt{\left(\frac{h^2}{a^2} + \frac{k^2}{b^2}\right)}$
$hk0$			$hk0$		
110	0.24	0.24	310	0.45	0.45
200	0.26	0.27	220	0.48	0.48
210	0.33	0.33	400	0.53	0.54
020	0.40	0.40	320	0.56	0.57
120	0.41	0.42	130	0.61	0.61

In addition to the random (001) orientation giving rise to figures 5 and 7 the crystals were often arranged to give the pattern of figure 6, especially in films formed on hot water. Here, the c -axis is again normal to the collodion, but the a - and b -axial directions are not randomly distributed parallel to the surface. Instead the rectangle formed by the a - and b -axes occurs in three positions only, in which corresponding axes make angles of 120° with each other. The angular extent of the arcs of figure 6 shows that the axes may occupy positions about $\pm 5^\circ$ on either side of the 120° directions. It then becomes possible for the (110) planes of one group to lie parallel to the (100) planes of one of the others, since the angle between these planes in the unit cell is 56° . The crystal grouping giving rise to figure 6, therefore, seems to be due to a form of twinning in which the (100) and (110) planes lie parallel to each other. This pseudo-hexagonal arrangement was first observed and explained by Garrido & Hengstenberg (1932), using the n -paraffin heneicosane, $C_{21}H_{44}$; it appears from the present work that the tendency to form pseudo-hexagonal groupings is common to all the normal paraffins. Similar diffraction patterns have also been obtained from stearic acid (Germer & Storks 1938) and from fatty acid esters (Coumoulos & Rideal 1941).

Occasionally, diffraction patterns due to large single crystals in (001) orientation have been observed, usually in conjunction with the pseudo-hexagonal type of pattern. They take the form of a rectangular grid of spots with unit translations $\lambda L/a$

and $\lambda L/b$. One film of triacontane deposited directly on to collodion gave the pattern of figure 8, due to crystals in (001) orientation and exhibiting strong alinement of the a - and b -axes to form a rough mosaic single crystal.

Summarizing the transmission results, the crystals are always in (001) orientation on the collodion. The distribution of the crystals about the normal may be random, pseudo-hexagonal, large single crystals, or small crystals so alined as to approximate to a single crystal.

THE HEATING OF NORMAL PARAFFIN FILMS

The experiments described above have been concerned primarily with the structure of normal paraffin films at room temperature, both before and after they have been subjected to shear. Another aspect requiring study from the lubrication viewpoint is the effect of heat upon such films. Accordingly, heated reflexion specimens have been examined.

The diffraction pattern on the fluorescent screen was watched as the temperature of the specimen increased, and it was observed that the bright parts of the pattern gradually became diffuse, and eventually disappeared. It was evident that the crystal structure broke down with increasing temperature, and finally even the vertical orientation of the molecular axes was lost, leaving an apparently blank screen. The loss of pattern was due simply to disorientation, and not to complete loss of the film by volatilization, because a coherent pattern always returned on allowing the specimen to cool.

The temperature at which disorientation occurred was taken to be that at which the screen finally became blank. It was measured for a range of thicknesses of the three paraffins used in the previous work, both on stainless steel and on copper with conventionally polished surfaces. The readings are plotted in figure 9 as functions of film thickness for the two metals. It must be remembered that the diffraction patterns give information applicable only to the top surfaces of the thicker films.

With all three paraffins the disorientation temperatures were low for the monolayers and increased with the thickness of the film, approaching asymptotically the bulk melting-points of 53, 68 and 76° C for the 24, 30 and 34 carbon atom chains, respectively.

The curves of figure 9 confirm that the method of film deposition employed gave rise to even continuous films, and not simply to crystal aggregates. Only on the former assumption can a dependence of disorientation temperature upon quantity deposited be explained. If the film had consisted of scattered crystal aggregates, the disorientation point should have been equal to the bulk melting-point, independently of the quantity deposited.

The heating of thick films of tetratriacontane has been described by Beeck, Givens & Smith (1941). They observed the fading of the electron diffraction pattern near the bulk melting-point, but believed it to be due to the evaporation

of the film. In view of the reappearance of the pattern on cooling, this view cannot be correct.

The significance of the disorientation temperature is of much interest from the lubrication viewpoint. In thick films it corresponds to the melting of the film; with monolayers, however, it may be regarded as a transition from a condensed to an expanded monolayer, using these terms in the same sense as when applied to monolayers on water. The lack of diffraction patterns from monolayers above the disorientation point shows that the molecular axes are randomly distributed in space. It follows that if the film is to remain on the surface, one end of each molecule must be firmly anchored, while the rest of the chain is free to take up any position in space above the surface, that is, the film is of the expanded type. The existence of such films has been visualized by Frewing (1943), as a result of friction experiments on long-chain compounds.

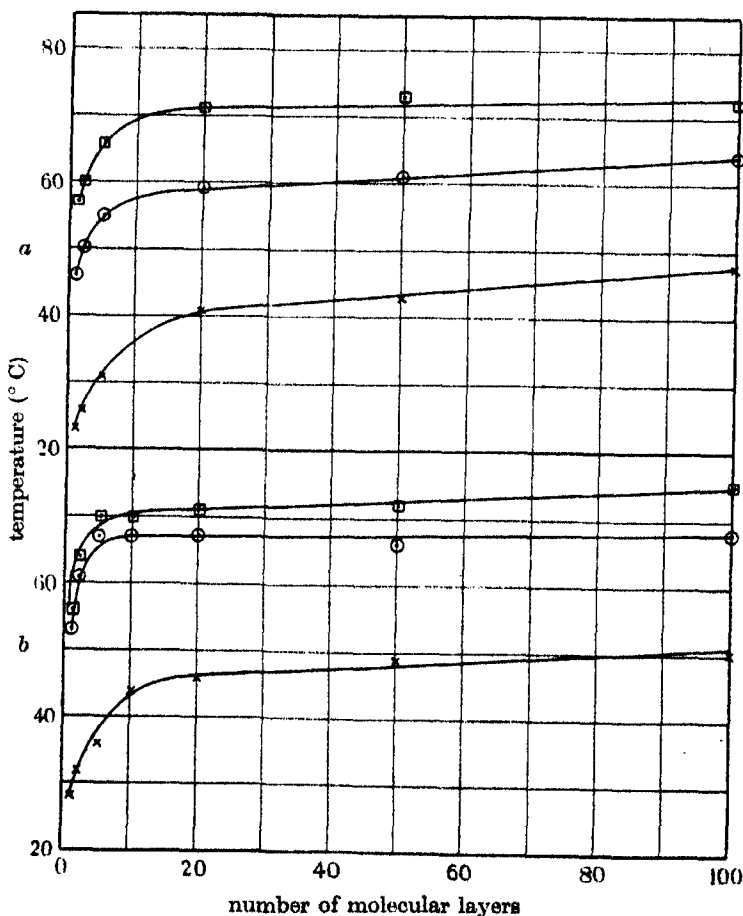


FIGURE 9. Disorientation temperatures as a function of film thickness. *a*, films on stainless steel; *b*, films on copper. × tetracosane. ○ triacontane. □ tetratriacontane.

DISCUSSION OF RESULTS

The above results show that a thin film of a n -paraffin upon a metal consists of a continuous and evenly spread layer, in which the molecular axes are normal to the substrate surface. The film bridges over valleys and pits upon the surface. In the case of monolayers, this means that the usual crystal structure is impossible, since the normal paraffin unit cell contains two pairs of molecules placed end to end. Nevertheless, the lateral grouping of the molecules on a rectangular cell of $7.45 \times 4.97 \text{ \AA}$ remains, although the crystal size is small. Thicker films are probably composed of even numbers of molecular layers arranged in the usual crystal structure. A film formed from a quantity of paraffin calculated to give an odd number, n , of layers is almost certain to consist of a mixture of films containing either $(n + 1)$ or $(n - 1)$ molecular layers.

Measurements have been made from the spacings of the diffuse layer lines of figure 1, plate 13, and of the almost vertical zones of figure 4, plate 13, of the value of the parameter a in the paraffin chain, that is, the distance between alternate carbon atoms. Correction was necessary in the former case for the inner potential effect and the value of a obtained as the result of measuring about 100 patterns was 2.66 \AA with a probable error of $\pm 0.03 \text{ \AA}$. The inner potentials observed were between 1 and 2 V, and were calculated as described by Thomson & Cochrane (1939). The value of $a = 2.66 \text{ \AA}$ is in excellent agreement with that obtained by Murison (1934) from electron diffraction patterns of fats and greases, but does not check with the accepted value, based on X-ray measurements, of 2.54 \AA . The figures for a calculated from the almost vertical zones of figure 4 were, however, in agreement with the X-ray value. The discrepancy between the electron and X-ray measurements, therefore, seems to be due to something inherent in either the production or measurement of horizontal layer lines of the type shown in figure 1. No adequate explanation has so far been found.

In correlating the reflexion and transmission results, it must be remembered that in reflexion experiments the area grazed by the beam is comparatively large; thus, the fact that in the experiments described above only random (001) orientation was observed by reflexion does not preclude the existence of large single crystals or of pseudo-hexagonal arrays at various azimuths to the beam within the area under examination, so that structures similar to those observed on collodion may well exist on metals.

In all the films examined, the molecular axes have been observed to make angles of either 90° or 3° , with the metal surface. The reason for the appearance of these angles becomes clearer from a consideration of the molecular arrangement in the (010) planes of the n -paraffin crystal; these planes are normal to the surface in the aligned (100) orientation with a 3° slope, as observed in the rubbed films. The molecular positions are shown diagrammatically in figure 10. The full lines represent molecules in the plane of the paper and the chain lines those at a distance $b/2$ above or below the paper. It is evident that the only planes normal to the

paper which can be drawn through this array without intersecting molecules are the (001) and (201) planes, whose traces have been inserted in figure 10. For a paraffin containing about thirty carbon atoms, the angles between these planes and the c -axis are 90° and about 3° respectively, which are indeed the only angles observed between the molecular axes and substrate. Hence, the occurrence of these angles in the films is to be expected simply on crystal structure considerations. The observation of other angles would mean that molecules were lying partly within and partly outside the surface. For this explanation to be valid, it is necessary to assume that parts of the surface are sufficiently flat to form a basis upon which (001) and (201) orientations can be initiated.

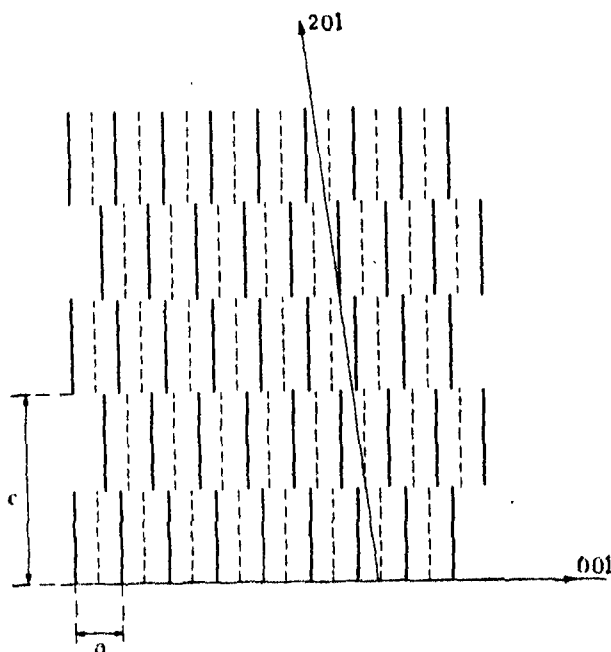


FIGURE 10. The 010 plane, n -paraffin crystal.

The above discussion offers some explanation of the (001) orientation and of the aligned (100) orientation with a 3° slope of the c -axis to the substrate and shows that the latter could more strictly be called (201) orientation. It is generally observed in conjunction with crystals in aligned (110) orientation, also with a 3° slope. The occurrence of (100) and (110) planes parallel to each other immediately suggests the existence of pseudo-hexagonal grouping, and, indeed, this furnishes an adequate explanation of the aligned (110) orientation. It is due to crystals built up on those in (201) orientation in accordance with the pseudo-hexagonal grouping.

When a thin film of a n -paraffin upon a metal surface is subjected to rubbing, the following sequence of events seems to take place. First, if the pressure applied

is sufficient, the thickness of the film is reduced mechanically until the surfaces are kept apart by an adsorbed layer of *n*-paraffin only a few molecules thick. Actual contact may occur between the surfaces at high spots. The work of Bowden & Tabor (1939) shows that such areas of contact are small, but if they occur, a large local rise of temperature is observed (Bowden & Ridler 1936). This temperature rise is momentary, but the heat liberated by successive contacts is sufficient to cause a rise in the general temperature of the surfaces.

Since the film of *n*-paraffin between the surfaces is only a few molecular layers thick, the increase in temperature is probably sufficient to cause disorientation of the film, so that the surfaces are actually kept apart by an expanded film. The molecules in such a film are free to take up any position required to support the load; they will remain between the surfaces unless conditions become so severe that desorption takes place. Should this occur, large scale contact between the surfaces is inevitable.

After passage of the rubbing surfaces over each other, the expanded film cools, and crystallization sets in. The foregoing discussion shows that the crystals should be in (001) and (201) orientations only, as observed; it would, however, be expected that the *c*-axes in the (201) orientation would be in random azimuth, and not in one direction only, parallel to the rubbing direction with the 3° tilt subtending the direction from which the rubbing surface approached. The explanation probably lies in the fact that the departing rubbing surface exerts a shear on the newly formed crystals. Those in (201) orientation are not as firmly attached to the surface as those in (001) orientation, because in the former case, only half of the molecules have their ends in contact with the surface, as is shown by figure 10. Also, the number of contacting molecules per unit area is much greater in the (001) orientation. Hence, the shearing force is much more likely to remove crystals in (201) orientation; those in the observed position are able to resist shear more effectively than those in any other azimuth, because there is in that case a component of the shearing force tending to hold the molecules on the surface.

The writer wishes to thank the Chairman of the Anglo-Iranian Oil Company for permission to publish this paper, and for providing facilities for the work. He would also like to express his gratitude to Professor G. I. Finch, F.R.S., for samples of pure normal paraffins, and for his continual interest and encouragement. Special thanks are also due to Mrs H. M. Cooke, who carried out a high proportion of the practical work involved.

REFERENCES

- Beeck, Givens & Smith 1941 *Proc. Roy. Soc. A*, **177**, 90.
 Bowden & Ridler 1936 *Proc. Roy. Soc. A*, **154**, 640.
 Bowden & Tabor 1939 *Proc. Roy. Soc. A*, **169**, 391.
 Charlesby, Finch & Wilman 1939 *Proc. Phys. Soc.* **51**, 479.
 Coumoulos & Rideal 1941 *Proc. Roy. Soc. A*, **178**, 421.

- Finch 1938 *J. Chem. Soc.* **141**, 1187.
Finch & Wilman 1937 *Ergebn. Exakt. Naturw.* **16**, 353.
Frewing 1943 *Proc. Roy. Soc. A*, **181**, 23.
Garrido & Hengstenberg 1932 *Z. Kristallogr.* **82**, 477.
Germer & Storks 1938 *J. Chem. Phys.* **6**, 280.
Jacquet & Rocquet 1939 *C.R. Acad. Sci., Paris*, **208**, 1012.
Muller 1928 *Proc. Roy. Soc. A*, **120**, 437.
Murison 1934 *Phil. Mag.* **17**, 201.
Thomson & Cochrane 1939 *Theory and practice of electron diffraction*, p. 155.
London: McMillan and Co., Ltd.

The anisotropy of thermal expansion as a cause of deformation in metals and alloys

BY W. BOAS AND R. W. K. HONEYCOMBE

*Council for Scientific and Industrial Research, Section of Tribophysics, Melbourne,
and Rosenhain Memorial Laboratory, University of Melbourne*

(Communicated by Sir David Rivett, F.R.S.—Received 24 January 1946)

[Plates 15-18]

Plastic deformation produced by repeated heating and cooling of certain non-cubic metals has been studied over a temperature range from -190 to 250°C . The extent of the deformation which results from the anisotropy of thermal expansion of individual crystals becomes greater as the temperature range is increased. In the case of cadmium, plastic deformation was detected even after 20 cycles over the temperature range 30 to 75°C ; in cyclic treatment to higher temperatures the deformation was associated with extensive grain-boundary migration. Cyclic treatment of tin, cadmium and zinc between room temperature and that of liquid air resulted in complex slipping and twinning, but grain-boundary migration was practically absent.

Cooling from the liquid state which represents half a thermal cycle sets up stresses in metals possessing anisotropy of thermal expansion, and leads to plastic deformation. This is shown by the detection of slip lines and grain-boundary accretion in certain cast non-cubic metals. In cadmium, subsequent annealing results in marked grain growth.

In order to investigate the interaction between the crystals of the two phases of a duplex alloy during cyclic thermal treatment, experiments were carried out with a series of tin-rich tin-antimony alloys. The majority of the alloys consisted of the tin-rich matrix in which particles of a hard second phase of cubic crystal structure were embedded. It was found that the deformation in the region of the boundaries between crystals of the two phases was considerably smaller than that in the region of the crystal boundaries of the anisotropic matrix. Similar results were obtained with tin-base bearing alloys.

1. INTRODUCTION

In a previous paper (Boas & Honeycombe 1946) it has been shown that certain non-cubic metals are deformed by heating and cooling. Electrolytically polished specimens of tin, cadmium and zinc were subjected to thermal cycles between 30 and 150°C , and in each case, signs of plastic deformation were detected after only a small number of cycles. The deformation in cadmium and zinc was clearly visible as slip, while in tin only distortion in the region of the grain boundaries was apparent. This deformation was shown to be due to the anisotropy of thermal expansion which is a characteristic of non-cubic metals. When lead, which is a cubic metal, was subjected to similar treatment, no signs of deformation were detected.

This paper describes an extension of the work to cover a wider range of experimental conditions and to include other metals and alloys; the effects of temperature, purity and preferred orientation on the deformation have been investigated. The work has led to a general consideration of the phenomenon in the casting and annealing of metals. The behaviour on cyclic thermal treatment of certain duplex alloys forming the basis of tin-base bearing metals has been examined in some detail.

The experimental technique employed was in most cases the same as that described previously, namely, the specimens of pure metals were annealed, electrolytically

polished, and subjected to cyclic thermal treatment. Microscopic examination could then be carried out satisfactorily without further polishing, for with most of the metals used little tarnishing occurred. Every care was taken to avoid mechanical deformation of the specimens. The electrolytic polishing method was found to be inadequate in dealing with the alloys, and normal metallographic polishing was resorted to, the specimens being annealed after final polishing.

2. DEFORMATION OF METALS BY CYCLIC THERMAL TREATMENT

(a) *Variation of maximum temperature of the cycle*

In the work described in the previous paper, the temperature range for the cyclic treatment of all the specimens was 30 to 150° C. Slip lines were, however, observed microscopically at temperatures far below the maximum temperature (150° C). Deformation would therefore be expected to occur during cyclic treatment within a smaller temperature range.

To confirm this point, a series of cadmium specimens, which had been annealed for 1 hr. at 200° C and were of similar grain size, was subjected to 20 cycles within various temperature ranges. Figure 1a, plate 15, shows an area representative of all specimens as polished. The specimens were immersed alternately in hot and cold oil-baths, the duration of a cycle being 7 min. Figure 1b-f, plate 15, show photomicrographs of typical areas from the specimens after cyclic treatment. For the range 30 to 75° C slip lines were produced after 20 cycles (figure 1b, plate 15), but no grain-boundary migration was observed in this temperature range even after 50 cycles. With increasing range of temperature, the deformation became much more pronounced, more grains were deformed, and grain-boundary migration became increasingly evident (figures 1c-e, plate 15). In the range 30 to 250° C after 20 cycles, the small original grains had been replaced by large grains. This process of grain growth resulted from the high maximum temperature of the cyclic treatment, and would have taken place in any case if the specimen had originally been annealed at 250° C instead of 200° C. The deformation was very pronounced, and was confined principally to the large new grains which themselves showed extensive grain-boundary migration. An examination of figure 1f, plate 15, shows the network of grain-boundary impressions produced by this migration; it can be seen that the slip lines cross the old grain boundaries. It should be emphasized that this specimen had originally a grain size similar to that of the other specimens of the series; some of the original grain-boundary impressions are still visible.

Marked grain growth was also observed during the cyclic treatment of spectroscopically pure tin (99.998 %) between 30 and 150° C (see figure 2a, plate 15). This had not been obtained with the less pure tin (99.85 %) used in the earlier experiments.

(b) *Cyclic cooling in liquid air*

The experiments described previously were carried out between 30° C and higher temperatures. If the deformation is due to the anisotropy of thermal expansion, then it should also occur when the cycles are carried out between room temperature

and lower temperatures. An advantage of such a temperature range is that grain-boundary migration is practically eliminated and recovery greatly reduced. To investigate this, the cycles were carried out by immersing the specimen, contained in a thin glass tube, in liquid air. To complete the cycle, the tube was removed and the specimen allowed to reach room temperature. The total time of such a cycle was 20 min.

After several such cycles, slip lines and some twinning were detected in zinc, cadmium and tin, but no effects were observed with lead. Some difficulty was encountered because of the tarnishing of the lead specimens, so to overcome this a lead specimen was sealed in a thin-walled glass tube in an atmosphere of dry oxygen-free nitrogen. To make a direct comparison a cadmium specimen was treated in the same manner, and the two specimens were subjected to 16 cycles. Subsequent microscopic examination showed slip lines and twinning in the cadmium, while the lead was free from any signs of deformation. No grain-boundary migration was observed in either case.

A specimen of spectroscopically pure tin (99.998 %) contained in a glass tube and subjected to 10 cycles showed slip lines and some twins. When a specimen of less pure tin (99.85 %) was similarly treated, few slip lines were observed, the main effect being grain-boundary accentuation. As shown in figure 2*b*, plate 15, the slip lines observed in very pure tin after cyclic treatment at low temperatures are straight. This is in contrast to slip lines produced during cyclic treatment at elevated temperatures, which are usually curved and less well defined (see figure 2*a*, plate 15). Similar curved slip lines have been described by Greenland (1937) in the case of mercury crystals.

It was also observed that the cyclic treatment at low temperatures produced complex slip indicated by the occurrence of two or more sets of slip lines in the one grain (see figure 2*b*). This phenomenon was observed in zinc, cadmium and tin.

(c) Specimens with preferred orientation

Deformation produced by cyclic thermal treatment has been shown to be due to the anisotropy of thermal expansion of the crystals comprising the material. If a preferred orientation exists in the metal, i.e. if the same crystallographic axes are parallel in all the crystals, then the thermal expansion parallel to the common boundary of neighbouring crystals will be the same, and no stresses will be set up on heating or cooling.

Many attempts were made to produce specimens with a strongly preferred orientation both by annealing subsequent to severe cold working and by casting. Such specimens should behave in a similar manner to a single crystal, and cyclic thermal treatment should produce no deformation. The results with pure metals (cadmium and zinc) were not conclusive because it was apparently not possible to produce a sufficiently strong preferred orientation. However, cast tin-antimony alloys possessing a marked columnar structure showed very little distortion in the

region of the grain boundaries even after 100 cycles between 30 and 150° C. This is in contrast to the behaviour of similar alloys consisting of equi-axed grains which, after the same treatment, showed marked deformation. This is shown in figure 3*a* and *b*, plate 16.

(*d*) *Aluminium and magnesium*

Lead has a cubic crystal structure with isotropic thermal properties. For this reason it shows no plastic deformation on heating or cooling. We should, therefore, expect that aluminium which is also cubic would behave similarly. Rolled and annealed aluminium (99.97 % pure) was electrolytically polished and subsequently annealed for 3 hr. at 250° C. It was then subjected to 50 cycles between 30 and 200° C, a greater temperature range than that used in most of the previous experiments. Micro-examination after this treatment revealed no signs of plastic deformation or grain-boundary migration.

Magnesium which has a melting-point close to that of aluminium and is not cubic was treated in exactly the same manner. Again, no deformation or grain-boundary migration was detected after 50 cycles. This is due to the fact that, although magnesium possesses a hexagonal crystal structure, its thermal properties are very nearly isotropic, and the stresses set up are presumably too small to produce observable effects.

3. CASTING AND ANNEALING OF METALS

(*a*) *Metals as cast*

The cooling of a metal from the liquid state already represents half a thermal cycle over a very wide temperature range. Consequently, if the metal possesses a high anisotropy of thermal expansion, then stresses will be set up in the cast metal.

Small ingots of zinc, cadmium, tin and lead were cast into preheated moulds with armour plate-glass bases. In the case of tin and lead it was possible to observe the free surface of the ingot, but with zinc and cadmium observations were more satisfactorily made on the under surface. In either case, no polishing of the resultant smooth surfaces was required in order to examine the ingots microscopically.

In cast cadmium and zinc, slip lines were observed in a number of grains. Typical examples are shown in figure 4*a* and *b*, plate 16. In all specimens examined, complicated networks of grain boundaries were apparent without etching. Similar networks have been observed in many metals by Vogel (1923). With tin, observation of the mirror-like free surface as the ingot solidified and cooled revealed, even to the naked eye, the gradual accentuation of the grain boundaries which had become very marked by the time room temperature was reached. This effect is similar to that observed previously during the cyclic thermal treatment of annealed tin specimens. In similar experiments with lead, no slip lines were produced, but grain boundaries were visible under the microscope, and a very limited amount of grain-boundary migration was detected.

(b) Annealing of cast metals

As the above experimental evidence indicates that stresses are present in some metals in the as-cast state, experiments were made to determine the effect of these stresses on subsequent annealing. Ingots of tin, cadmium, zinc and lead, $1\frac{1}{2} \times 3 \times \frac{1}{4}$ in., were cast in a shallow steel mould, which was, in each case, preheated before casting.

Cadmium and lead ingots cast into moulds preheated to approximately 300° C were macro-etched on the free surfaces and photographed. The ingots were then annealed in an air oven for 24 hr. at 290 to 300° C; they were allowed to cool slowly and were then re-etched. In each experiment, a lead ingot and a cadmium ingot were treated simultaneously to ensure comparable results. Figure 5 *a* and *b*, plate 17, are photomicrographs of ingots prior to annealing, while figure 5 *c* and *d*, plate 17, show the same two ingots on re-etching after annealing. It is apparent that marked grain growth has occurred in the cadmium ingot, while the lead ingot has remained substantially unchanged. This effect was observed in castings over a range of grain sizes (0.5–3 mm.) and occurred both when the specimens were placed directly into the air oven at 290° C and when they were slowly heated up with the oven to the final temperature.

In tin ingots (99.975 % pure), annealed for 24 hr. at 190 to 200° C, grain growth was detected in one case only, and even here it occurred in a limited area of the ingot. This would be expected in view of the relatively small anisotropy of thermal expansion in tin.

Zinc ingots (99.99 % + pure) of the same size were cast and annealed for 24 hr. at 365° C. No general change in grain size was observed by the macroscopic technique employed, even in the case of an ingot annealed for 96 hr. at 380 to 390° C, although some localized absorption of grains occurred. This result is surprising in view of the great anisotropy of thermal expansion in zinc. Further experiments with specimens of equally pure zinc, but possessing different grain sizes, were carried out and the same results obtained. The absence of grain growth may be a question of purity, but no definite explanation can be offered.

4. DEPTH AND EXTENT OF DEFORMATION

The metallographic observation of plastic deformation, as indicated by slip lines, twinning and distortion in the vicinity of grain boundaries, is limited to the surface of the metal. Removal of the surface layer by etching or polishing removes also the slip lines; however, X-ray photographs taken after etching will reveal if any deformation is present in the interior of the specimen.

A zinc specimen approximately 0.15 in. thick was subjected to 100 cycles between 30 and 150° C and an X-ray back-reflexion photograph was taken; further back reflexion photographs were then made after the removal of 0.005, 0.012 and 0.080 in. from one face of the specimen by etching. In each case, lattice distortions were shown by the blurring of the spots, and no marked difference in the extent of blurring was detected. Thus the deformation occurs throughout the whole specimen and is not restricted to the surface layers.

It might thus be expected that sufficiently prolonged cyclic thermal treatment would produce a change in the physical properties of the metals. To investigate this, hardness measurements were carried out on specimens of cadmium and zinc after various numbers of cycles. Even after 400 cycles between 30 and 150° C, practically no increase in hardness was detected in either zinc or cadmium. Since it was possible that recovery occurred during the thermal cycles with the elimination of the work hardening, cyclic treatment of zinc was carried out between room temperature and the temperature of liquid air. After 75 such cycles an increase in hardness of only 2 Brinell units was obtained. This increase indicates that, in spite of the marked visual evidence, the plastic deformation is not very severe. This result can be compared with the increase in hardness produced by mechanical working. A specimen of annealed zinc compressed by 9 % showed an increase in hardness of 5 Brinell units.

5. PURITY OF METALS

It has already been mentioned that a significant difference has been observed in the behaviour of two grades of tin when subjected to cyclic treatment. First, very few slip lines were observed in specimens of the less pure tin (99.85 %), both when the treatment was carried out in liquid air and at elevated temperatures. The more obvious sign of deformation was distortion occurring in the region of the grain boundaries. When specimens of spectroscopically pure tin (99.998 %) were employed, many slip lines were detected as a result of cyclic treatment in both temperature ranges (figures 2 *a* and *b*, plate 15). Another point of difference between the two materials was the extent of grain-boundary movement at elevated temperatures. In the less pure tin, grain-boundary migration occurred during cyclic treatment between 30 and 150° C. A specimen of spectroscopically pure tin, when subjected to 25 cycles in the same temperature range, showed pronounced grain growth. This occurred early in the cyclic treatment because marked slip lines were evident in the large new grains, and only very slight evidence of slip lines could be detected in the small original grains. The large grains once formed were not stable, but showed grain-boundary migration as the cyclic treatment was continued.

The effect of purity was also marked in the case of zinc. Zinc of 99.95 % purity* showed no grain-boundary migration after cyclic thermal treatment in the range 30 to 150° C, whereas a purer grade of zinc (99.99 % +)* showed grain-boundary

* The results of spectroscopic analyses carried out on specimens prepared from the two grades of zinc were as follows:

	grade 1	grade 2
lead %	0.0025	0.03
iron %	0.004	0.02
cadmium %	c. 0.002	c. 0.002
copper	trace	trace
silver	faint trace	very faint trace
estimated purity %	99.99 +	99.95

No indications of other impurities were found.

migration associated with the usual signs of deformation. Figure 6a, plate 17 shows an area on such a zinc specimen after 50 cycles between 30 and 150° C. Some thickening of the grain boundaries is evident; at higher magnifications this can be resolved into a series of grain boundaries (figure 6b, plate 17).

These results are summarized in table 1.

TABLE 1. EFFECT OF IMPURITIES IN TIN AND ZINC

metal	temperature range (° C)	deformation	boundary movement
Sn 99.85 %	30-150	few slip lines accentuation of grain boundaries	some migration
99.998 %	30-150	many curved slip lines	marked grain growth
Zn 99.95 %	30-150	straight slip lines	no migration
99.99 % +	30-150		some migration
Sn 99.85 %	-180-30	few slip lines accentuation of grain boundaries	none
99.998 %	-180-30	straight slip lines multiple slip, some twinning	none

6. EXPERIMENTS WITH DUPLEX ALLOYS

The deformation described above occurs as a result of interaction between crystals of the one phase. In alloys in which two phases are present, deformation due to the different thermal expansion of the two phases is superimposed on the deformation due to the anisotropy of the primary phase. In order to investigate the relative importance of these two effects some experiments were carried out with tin-antimony alloys. These alloys are of some practical importance as they form the basis of tin-base bearing alloys.

Tin-rich tin-antimony alloys were chosen because large particles of a second phase could be obtained as a suitable distribution in the primary solid solution. This solid solution has the tetragonal structure of tin, while the second phase is the inter-metallic compound SnSb which possesses a cubic crystal structure (Bowen & Morris-Jones 1931). The microstructure of the 5 % antimony alloy consists solely of the solid solution, whereas in the alloys of higher antimony content, cuboids (SnSb) are embedded in the tetragonal matrix (Hanson & Pell-Walpole 1936).

Alloys containing approximately 5, 10, 15 and 20 % antimony were cast and small specimens were mechanically polished. After final polishing, the specimens were annealed for 1 hr. at 200° C, and were then subjected to cyclic thermal treatment between 30 and 150° C in glass tubes in order to minimize tarnishing. Selected areas were photographed at intervals during the experiment. In all cases the surfaces of the specimens acquired a 'rumpled' appearance as a result of distortion in the vicinity of the grain boundaries of the matrix.

That these distortions coincide with the grain boundaries has been shown by macro-etching. The deformation has the appearance of cracks in the alloy. How-

ever, it has been shown by cross-sectioning that the cracks are usually limited to the surface of the metal, although the deformation extends much farther. With increasing antimony content, this surface rumpling became less pronounced. A few slip lines were detected in the 5 % antimony alloy (figures 7 *a* and *b*, plate 18), but none were observed in the other alloys (figures 7 *c* and *d*, plate 18), although the matrix has nearly the same composition in all cases. These observations indicate that the cuboids have a stiffening effect on the matrix, which shows less plastic deformation on cyclic thermal treatment the greater the number of cuboids present. Two factors contribute to this stiffening of the solid solution comprising the matrix: first, the cuboids are much harder than the matrix, and secondly, their presence in increasing numbers reduces the volume of the anisotropic phase in the alloy. The effect is clearly shown in figures 7 *a-d*, plate 18. Micro-examination of the cuboids embedded in the grains of the matrix showed slight distortion in the vicinity of some cuboids, but the main deformation was restricted to the region of the grain boundaries of the primary solid solution.

Similar results were obtained with tin-base bearing alloys. The composition typical of those investigated is as follows:

tin	91.9 %	copper	4.0 %
antimony	4.0 %	nickel	0.1 %

In such alloys the second phase consists of acicular particles of CuSn (Farnham 1934) which possess a hexagonal crystal structure (Westgren & Phragmén 1928). Again on cyclic thermal treatment, deformation occurred as distortions in the regions of the grain boundaries of the matrix. The hard particles contributed little to the total deformation. No such distortions were obtained with lead-base bearing alloys.

7. DISCUSSION AND CONCLUSIONS

The work described in our previous paper was based on experiments carried out with tin, cadmium, zinc and lead in the temperature range 30 to 150° C. This work has now been extended by varying the temperature range and employing additional metals. The deformation observed in non-cubic metals is due to their anisotropy of thermal expansion, and a theoretical estimation of the stresses shows that their magnitude is directly proportional to the temperature range. The extent of plastic deformation should therefore decrease as the temperature range of the cycle is narrowed. This has now been shown qualitatively to be the case. The high sensitivity of cadmium to deformation on cyclic thermal treatment should be noted, variation of the temperature between 30 and 75° C being sufficient to produce slip lines.

The fact that the slip lines are not due to the relief of internal stresses present in the metal as a result of previous treatment has already been shown. This is further supported by the following two observations. First, during the cyclic treatment of cadmium between 30 and 250° C, and tin between 30 and 150° C, slip lines appear in new crystals formed as a result of marked grain growth early in the cyclic treatment.

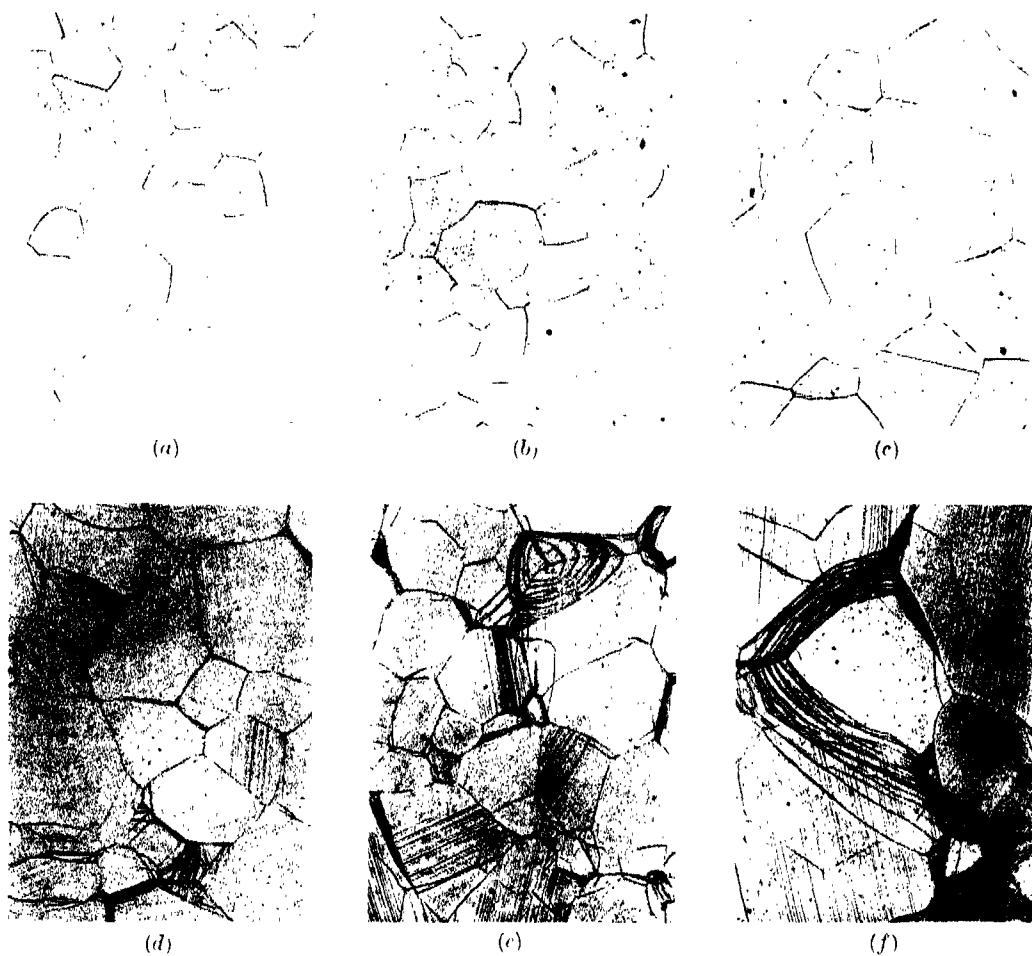


FIGURE 1

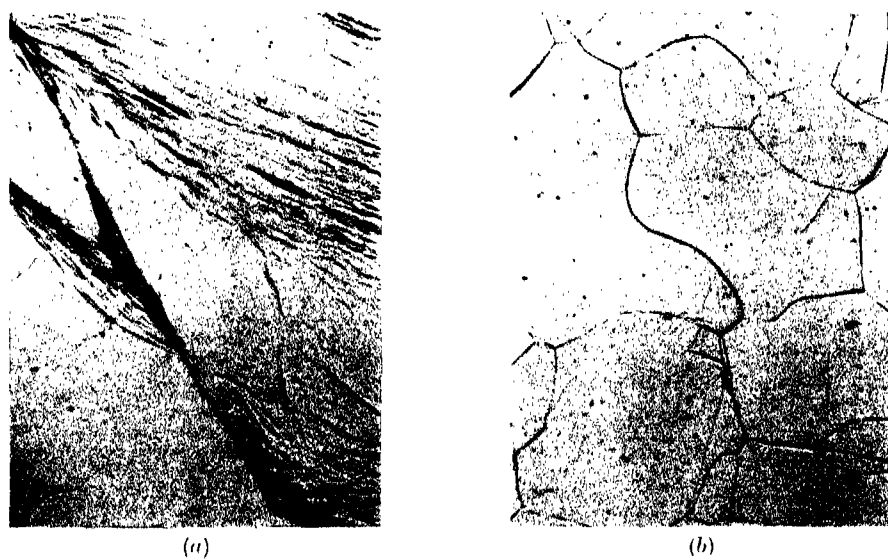
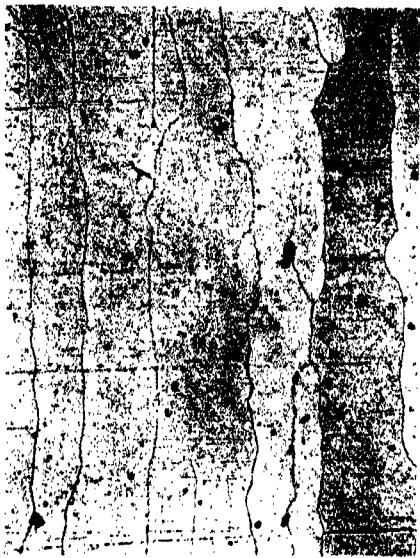


FIGURE 2

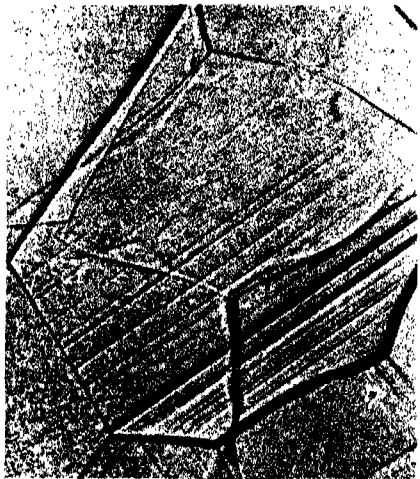


(a)

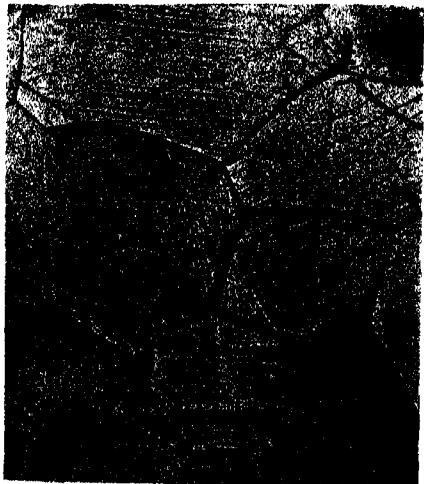


(b)

FIGURE 3



(a)



(b)

FIGURE 4

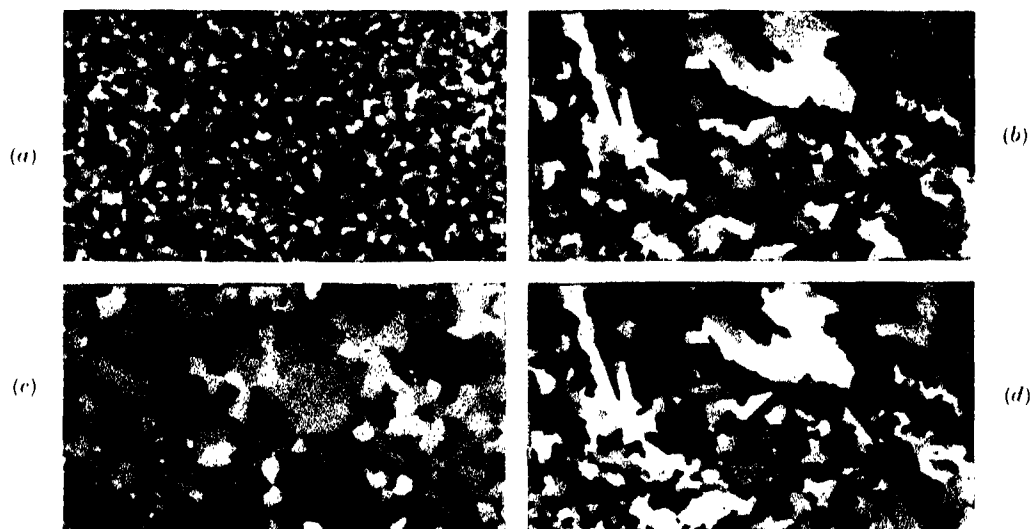


FIGURE 5

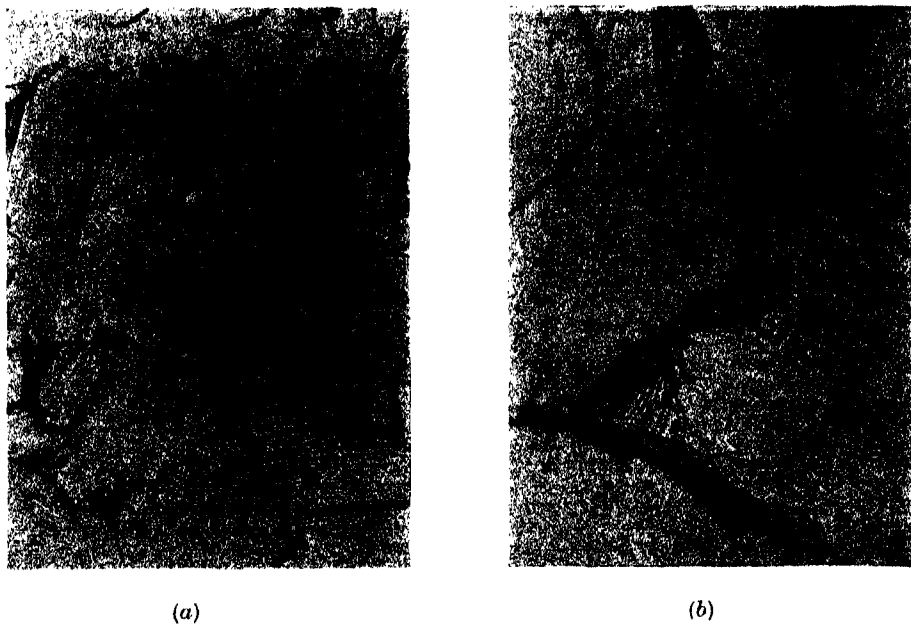


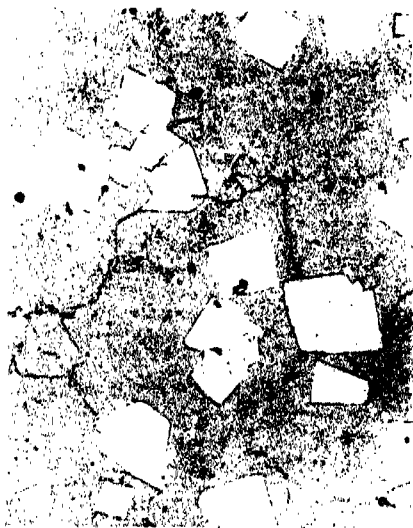
FIGURE 6



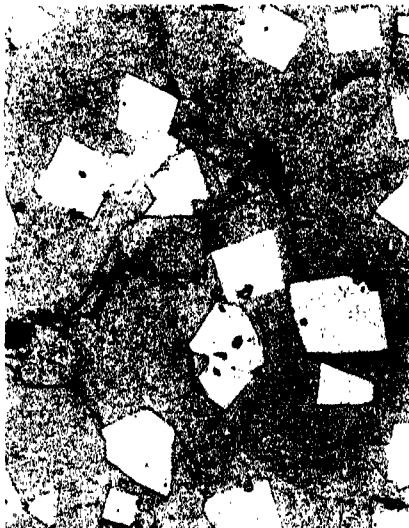
(a)



(b)



(c)



(d)

FIGURE 7

(figures 1 and 2a, plate 15). Secondly, zinc, cadmium and tin are plastically deformed when subjected to cyclic treatment between room temperature and the temperature of liquid air. Under these conditions the relief of pre-existing internal stresses in metals by slip cannot be expected, and consequently the deformation must have another source.

When the experiments were carried out in liquid air, the density of slip lines was similar to that obtained during cyclic treatment to elevated temperatures, as far as could be estimated by microscopic examination. The stress is apparently more complicated, as it gives rise to complex slip and some twinning in zinc, cadmium and tin. The temperature range employed (approximately 210°C) was greater than that used in the previous work (120°C); on the other hand, the critical shear stress of these metals at the temperature of liquid air is at least twice the value at 150°C (Schmid & Boas 1935). Thus a higher stress can be maintained than at more elevated temperatures without being relieved by plastic deformation. It follows that the change in temperature, and not the direction of this change, is the important factor in the deformation of non-cubic metals as a result of the anisotropy of thermal expansion. The cubic metal lead showed no signs of deformation when given cyclic treatment both to low temperatures and elevated temperatures.

The metals zinc, cadmium, tin, and lead, which were used in the majority of the experiments, possess low melting-points. In the case of cadmium and lead, the comparison of the behaviour on cyclic thermal treatment was particularly appropriate because of the closeness of their melting-points. The work was then extended to aluminium and magnesium. These metals also possess similar melting-points which are, however, considerably higher than those of the other four metals. As would be expected the cubic metal aluminium showed no signs of plastic deformation even after 50 cycles between 30 and 200°C .

After the same treatment no sign of plastic deformation could be observed in magnesium. In the appendix to our previous paper, a calculation was made of a factor indicating the comparative likelihood of slip for the various metals. This factor was 2 for tin, 12 for cadmium and 15 for zinc. A similar calculation for magnesium, in which the maximum coefficient of thermal expansion is only 6 % higher than the minimum, gave a figure of 0.04.

Grain growth has been shown to occur when cast cadmium is annealed (Cook 1923), and the observations on cadmium described above confirm this result. The same effect has been observed during the annealing of cast zinc of high purity (Guillet 1943), although the present authors have been unable to confirm this result. As early as 1923 Desch suggested that this grain growth occurs because of stresses set up as a result of the anisotropy of thermal expansion during cooling from the liquid state. The view that deformation is a prerequisite for grain growth is implicit in this explanation. Desch mentioned the absence of grain growth during the annealing of a cast cubic metal, namely, gold (Fraenkel 1922), in support of his contention. This has been further substantiated by the present authors' observations on lead. Recently, László (1944) has calculated the order of magnitude of such

stresses; he has suggested that recrystallization occurs during cooling of non-cubic metals and produces a preferred orientation in the material resulting in lower strain energy.

The detection of slip lines on the surface of cast cadmium and zinc which has already been described (figures 4 *a* and *b*, plate 16) fully supports the above explanation of grain growth. It therefore appears impossible to obtain these non-cubic metals in a completely strain-free condition at room temperature. The stresses may be relieved by annealing at sufficiently high temperatures, but new stresses will be produced on cooling again to room temperature, resulting in further plastic deformation. This is in direct contrast to the conventional principle that stresses due to plastic deformation are removed by prolonged treatment at elevated temperatures. Thus in determining a suitable heat treatment to obtain non-cubic metals in a strain-free condition, it is doubtful whether a high-temperature annealing, after which high new stresses are produced on cooling, is preferable to a low-temperature annealing in which the previous stresses are incompletely eliminated, but only slight new stresses are produced. The most satisfactory procedure probably consists of a low-temperature annealing over a long period of time.

Grain-boundary migration in zinc was only detected with metal of very high purity (99.99 % +) and not in a less pure grade (99.95 %) when specimens were subjected to cyclic thermal treatment between 30 and 150° C. This emphasizes again the familiar effect of insoluble impurities on the grain-boundary movement in metals. The main impurities in the zinc were lead and iron. The solubility of lead in zinc is less than 0.03 % (Peirce 1922), and the solubility of iron in zinc is no more than 0.0028 % (Truesdale, Wilcox & Rodda 1936). Thus in the zinc of 99.95 % purity which did not show grain-boundary migration, at least some iron will be present as a second phase. On the other hand, grain-boundary migration does occur in the purer zinc where these impurities are held almost entirely in solid solution.

The nature of the deformation produced by cyclic thermal treatment of tin-antimony alloys was similar to that obtained in the case of pure tin, namely, principally distortion at the grain boundaries of the primary solid solution.

Grain-boundary migration, which occurred to a marked degree in pure tin, was absent in the alloys of higher antimony content. However, a limited amount of migration occurred in the 5 % antimony alloy. Carpenter & Elam (1920) also observed grain-boundary migration and accentuation in a 1.5 % antimony alloy in experiments of a similar nature designed to investigate the process of grain growth during annealing.

It would be expected that some interaction should occur between the cuboids and the matrix in the tin-antimony alloys of higher antimony content, for at least in certain directions there must exist a difference in thermal expansion. Some evidence of deformation as a result of this has been observed, but the effect is very much smaller than that occurring at the grain boundaries of the matrix. This small effect can be understood because the value of the coefficient of thermal expansion of the phase SnSb lies between the limits of the coefficients of expansion of the

tin-rich matrix.* In tin-base bearing alloys only slight interaction has been observed between the matrix and the acicular particles (CuSn) which possess a hexagonal structure. The main deformation again occurs in the matrix. The same observation has been made in the case of lead-base bearing alloys where there was no interaction between the hard particles of cubic crystal structure (CuSn) and the cubic lead-rich matrix. In these alloys, however, no deformation is observed in the matrix because of its isotropic nature.

It is clear that in the tin-base alloys, the surface cracks occur in the vicinity of the grain boundaries of the tin-rich solid solution, and correspond to the grain-boundary accentuation observed in pure tin. In the case of bearings, these surface irregularities can extend deeply into the bearing alloy lining, and may ultimately reach the steel outer shell. As these distortions occur in the region of the grain boundaries, their distribution will be related to the size and shape of the grains of the tin-rich phase. The extent to which the effect produces defects in tin-base bearings under service conditions will depend primarily on two factors, namely, the operating temperature of the bearing and the number of times it is raised to this temperature. The above experiments have indicated that marked defects occur even after a few cycles in a temperature range over which many bearings operate.

This work has been confined to the effect of cyclic temperature changes on metals and alloys of relatively low melting-point. Of the metals of higher melting-point, few possess a non-cubic crystal structure, and no determinations of the anisotropy of thermal expansion have been made. The only information refers to cobalt, in which case no anisotropy of thermal expansion has been detected (Wassermann 1932).

In many other substances which possess marked anisotropy of thermal expansion, stresses will occur when they are subjected to temperature changes. For example, the existence of such stresses in rocks had already been discussed by Howe (1910). He considered that the weathering of rocks is partially due to stresses produced by temperature variations, and distinguished two possibilities by which minute intergranular thrusting is produced. Stresses are set up as a result of differences in the coefficients of thermal expansion, first of the various constituents of the rock, and secondly within grains of the one constituent as a result of the anisotropy of thermal expansion. Lord Rayleigh (1934) found that after marble had been heated to a temperature of 100° C its rigidity was diminished. This effect became increasingly marked as the temperature was raised further. He interpreted the results as due to the anisotropy of thermal expansion of calcite. As a result of the stresses set up during heating, the structure is dislocated and loosened with increase of volume, the process being irreversible. It is thus evident that the continual heating and

* A determination of the coefficient for SnSb gave the value 17×10^{-6} . Maximum and minimum values for tin are 30.5 and 15.5×10^{-6} , and these will not be materially altered by the presence of antimony in solid solution. Thus the maximum difference in thermal expansion between hard particles and matrix is smaller than that possible between two matrix grains.

cooling of rocks, with the resultant inhomogeneous expansion and contraction, contribute to the ultimate disintegration which is apparent both in stone buildings and in the landscape.

The work described in this paper was carried out as part of the programme of the Section of Tribophysics of the Council for Scientific and Industrial Research, Australia.

We wish to express our thanks to Dr F. P. Bowden and Dr D. Tabor for their interest and encouragement in the work, and to Dr S. H. Bastow and Professor J. Neill Greenwood for very helpful discussions. We also wish to thank the staff of the Munitions Supply Laboratories, Maribyrnong, for carrying out the spectroscopic analyses. Finally, we are grateful to Professor E. J. Hartung for laboratory facilities in the Chemistry School of the University of Melbourne.

REFERENCES

- Boas, W. & Honeycombe, R. W. K. 1946 *Proc. Roy. Soc. A*, **186**, 57-71.
 Bowen, E. G. & Morris-Jones, W. 1931 *Phil. Mag.* **12**, 441-462.
 Carpenter, H. C. H. & Elam, C. F. 1920 *J. Inst. Met.* **24**, 83-131.
 Cook, M. 1923 *Trans. Faraday Soc.* **19**, 43-48.
 Desch, C. H. 1923 *Trans. Faraday Soc.* (discussion), pp. 48-49.
 Farnham, G. S. 1934 *J. Inst. Met.* **55**, 69-70.
 Fraenkel, W. 1922 *Z. anorg. Chem.* **122**, 295-298.
 Greenland, K. M. 1937 *Proc. Roy. Soc. A*, **163**, 28-34.
 Guillet, L. (Jr.) 1943 *C.R. Acad. Sci., Paris*, **216**, 642-644.
 Hanson, D. & Pell-Walpole, W. T. 1936 *J. Inst. Met.* **56**, 299-308.
 Howe, J. A. 1910 *The geology of building stones*, pp. 342-343. London: Arnold.
 László, F. 1944 *J. Iron Steel Inst.* (pre-print).
 Peirce, W. M. 1922 *Trans. Amer. Inst. Min. (Metall.) Engrs*, **68**, 768-769.
 Rayleigh, Lord 1934 *Proc. Roy. Soc. A*, **144**, 266-279.
 Schmid, E. & Boas, W. 1935 *Kristallplastizität*. Berlin: Springer.
 Truesdale, E. C., Wilcox, R. L. & Rodda, J. L. 1936 *Trans. Amer. Inst. Min. (Metall.) Engrs*, **122**, 192-228.
 Vogel, R. 1923 *Z. anorg. Chem.* **126**, 1-38.
 Wassermann, G. 1932 *Metallwirtschaft*, **11**, 61-65.
 Westgren, A. & Phragmén, G. 1928 *Z. anorg. Chem.* **175**, 80-89.

DESCRIPTION OF PLATES 15-18

Plate 15

FIGURE 1. Typical areas on cadmium specimens after 20 cycles over various temperature ranges. (a) as polished, (b) 30-75° C, (c) 30-100° C, (d) 30-150° C, (e) 30-200° C, (f) 30-250° C. ($\times 60$.)

FIGURE 2. Areas on specimens of spectroscopically pure tin. (a) after 25 cycles between 30 and 150° C, showing marked grain growth and curved slip lines ($\times 40$), (b) after 10 cycles by direct immersion in liquid air, showing straight slip lines and twinning ($\times 80$).

Plate 16

FIGURE 3. The effect of crystal orientation on the deformation produced by cyclic thermal treatment. Typical areas on tin-antimony specimens containing 10 % antimony after 100 cycles between 30 and 150° C. (a) specimen with columnar crystals (preferred orientation) ($\times 44$), (b) specimen with equi-axed crystals (random orientation) ($\times 80$).

FIGURE 4. Plastic deformation in cast metals as shown by slip lines on the smooth under surfaces of ingots. (a) cadmium ($\times 160$), (b) zinc ($\times 40$).

Plate 17

FIGURE 5. The annealing of cast metals. (a) cadmium ingot—as cast, (b) lead ingot—as cast, (c) same as (a) after annealing, showing marked grain growth, (d) same as (b) after annealing, showing no grain growth. (All macrographs $\frac{1}{2}$ natural size.)

FIGURE 6. Cyclic thermal treatment of very pure zinc between 30 and 150° C. (a) typical area after 50 cycles showing slip lines and thickening of grain boundaries ($\times 80$), (b) an area at higher magnification ($\times 400$). The thickening is resolved into a grain-boundary network.

Plate 18

FIGURE 7. Cyclic thermal treatment of tin-antimony alloys between 30 and 150° C. (a) area on a 5 % antimony alloy after 10 cycles, (b) same area as (a) after 100 cycles, (c) area on a 15 % antimony alloy after 10 cycles, (d) same area as (c) after 100 cycles. ($\times 80$).

The two-dimensional hydrodynamical theory of moving aerofoils. IV

BY ROSA M. MORRIS, PH.D.

(Communicated by G. Temple, F.R.S.—Received 21 March 1946)

In the first part of this paper opportunity has been taken to make some adjustments in certain general formulae of previous papers, the necessity for which appeared in discussions with other workers on this subject.

The general results thus amended are then applied to a general discussion of the stability problem including the effect of the trailing wake which was deliberately excluded in the previous paper. The general conclusion is that to a first approximation the wake, as usually assumed, has little or no effect on the reality of the roots of the period equation, but that it may introduce instability of the oscillations, if the centre of gravity of the element is not sufficiently far forward. During the discussion contact is made with certain partial results recently obtained by von Karman and Sears, which are shown to be particular cases of the general formulae.

An Appendix is also added containing certain results on the motion of a vortex behind a moving cylinder, which were obtained to justify certain of the assumptions underlying the trail theory.

INTRODUCTION

1. The previous papers of this series (Morris 1937, 1938, 1939; referred to subsequently as I, II and III) have dealt in detail with the two-dimensional potential motion round an infinite cylinder of quite general shape. In I, I derived the kinetic

energy of the fluid motion when the circulation is zero, and in the more general case the forces and couple on the cylinder. The theory was extended in II to include the effect of a trail of vortices extending downstream behind the aerofoil from an assumed trailing edge. In the third paper the general formulae obtained in I and II have been reduced for the special case of the Joukowski aerofoil. The lift and moment for the rectilinear motion was then determined and was followed by a discussion of the stability of this motion by consideration of the dynamics of a small oscillation about it.

In a brief discussion at the end of II, when the general formulae obtained were reduced for special cases of certain cylinders which had been treated by other writers, I pointed out that the results did not agree in certain small details with those of von Karman & Burgers in Durand (1935).

Since the publication of these papers I have had the privilege of discussing certain aspects of the work and the results in these papers, with other writers on the subject. After arguing certain aspects of the trail theory as presented in II with Dr A. E. Green, I received a note from Dr von Karman and Dr Sears pointing to an error in the interpretation of the general results in paragraph 4.3 of this same paper. On correcting this interpretation another inconsistency in the results appeared due to an improper use of the transformation theory. On adjustment of these points, to which the first paragraphs of the present paper are devoted, slightly different results are obtained which now show complete agreement with the results of von Karman & Burgers and others for the special cases with which they deal. I have to thank Dr Green and also Dr von Karman and Dr Sears for their encouraging interest in the work and for the help they have given in correcting the errors inherent in the early publications.

Since II was published, a paper by von Karman & Sears (1938) has appeared dealing with the trail theory results of Wagner, Glauert, Burgers and others. In that paper the authors, adhering to the bound vortex theory of one of them, estimate the effect of the trail in certain oscillatory motions of the aerofoil. The theory of bound vortices involves the introduction of many additional hypotheses relating to the distribution and motion of the vortices, and although it may be a convenient way of forming the linearized equations of motion, it seems desirable to reformulate the discussion on the simpler physical lines of the present theory, using exact equations, and involving no assumptions beyond those involved in the underlying hydrodynamical theory, and the existence of circulation and vortices satisfying certain necessary conditions. This discussion is carried out in paragraph 6 but in a less restricted form than usual, so that it can be applied to the derivation of a period equation for the oscillations obtained in paragraph 2, from which it is possible to derive conditions for stability.

The paper closes with a discussion of certain relevant special cases of general results obtained in II, formulated with the object of estimating the legitimacy of certain assumptions underlying the trail theory, more particularly those relating to the velocity and distribution of the trail vorticity.

Before concluding this introduction I should like to refer here to more recent papers by Sohngen (1940), Sears (1940, 1941), Kussner & Schwarz (1940), Schwabe (1943), and von Karman & Tsien (1945), none of which, so far as I can gather, overlap the developments of the present paper, being rather extensions of the special linearized theories of previous writers. The present paper was, however, completed in 1941, publication having been delayed for obvious reasons, and proper reference to the work in these later papers will be given in connexion with further work which has been planned.

THE GENERAL FORMULAE

2. It was pointed out in II that a change in sign should be introduced into the results of I owing to the convention that $\eta = +\infty$ corresponds to $z = +\infty$. Now notice, in addition, that since there is a discontinuity in the motion at the trailing edge the contour of the cylinder does not lie completely in a region of irrotational motion. Therefore Bernoulli's equation for the pressure at all points of the contour cannot be used. To avoid the difficulty it must be assumed that the trailing wake is confined to a narrow band from the trailing edge, and then provided the contour integrals are taken from the trailing edge round the contour and back to the trailing edge, the previous analysis holds. This means that in the ζ -plane the limits instead of being 2π and 0 should be $+\pi$ and $-\pi$. In effect this only means a change in the value of one term in the expression for the forces, and one term in the expression for the couple, viz. the \dot{K} term, which appears now as

$$\rho \dot{K} c \left\{ \bar{a}_0 + \sum_{n=2}^{\infty} (-)^n \bar{a}_n \right\}$$

in the expression for the force, and

$$-\frac{1}{2} \rho \dot{K} c^2 \left\{ \sum_{n=1}^{\infty} (-)^n (b_n + \bar{b}_n) \right\}$$

in the expression for the couple.

It will be noticed that although in I no mention is made of a trailing wake, the fact that a variable circulation is assumed must I think indicate that there is a point of discontinuity on the contour which must be taken at the trailing edge, although the trailing wake is neglected.

The results from I for the force and couple used in II are derived from a potential function

$$\Omega = \Omega_a - \frac{K\zeta}{2\pi},$$

and are taken in the form

$$\begin{aligned} Y + iX &= -\pi\rho i \{ A c^2 \bar{w} - 2\bar{B} c^2 w - i\omega c^3 \bar{D} \} \\ &\quad - \pi\rho\omega \{ A c^2 \bar{w} - 2\bar{B} c^2 w - i\omega c^3 \bar{D} \} \\ &\quad - \rho i c \omega K a_1 + \rho \bar{w} K + \rho \dot{K} c \left\{ \bar{a}_0 + \sum_{n=2}^{\infty} (-)^n \bar{a}_n \right\}, \end{aligned}$$

and Γ as the real part of

$$-\pi\rho Cc^4\dot{w}-\pi\rho ic^3D\dot{w}-\pi\rho iw\{Ac^2\bar{w}-2\bar{B}c^2w-iwc^3\bar{D}\} \\ -\frac{1}{2}K\rho c^2\sum_{n=1}^{\infty}(-)^n(b_n+\bar{b}_n)+\rho c\bar{w}Ka_1-\rho ic^3Kb_0.$$

It is these results in their original form that I have used in III to find the forces and couple on the special Joukowski aerofoil. The result of the above change in the K term will mean that some of the expressions in III are incorrect, as well as the coefficients in the equations (9.1) and (9.2) for λ . With the approximations I have made, however, the final condition (9.5) for the stability of the flat plate, remains unchanged, whilst that for the Joukowski aerofoil becomes

$$I_G n^2/U^2 > 4\pi\rho\{a(a-2x_0-x_G)+y_0y_G\}\cos 2\alpha+4\pi\rho\{ay_G+y_0x_G\}\sin 2\alpha.$$

It can be easily seen that the conclusions arrived at in that paper still hold.

3. Again in calculating in II the acceleration terms resulting from the variations in the positions and strengths of the trailing vortices, the results were deduced by summation from those for a single vortex, without allowing for the fact that the number of vortices in the trail is continuously increasing. The correct method is to consider the complete integral, the limits of which, since the axes are moving axes, are functions of the time. The acceleration terms in the integral for the forces, for example, are

$$-\frac{1}{2}\rho\int_c\left\{\frac{\partial\Omega}{\partial t}+\frac{\partial\bar{\Omega}}{\partial t}\right\}d\bar{z},$$

and the terms being all superposable, the part corresponding to the wake is

$$-\frac{1}{2}\rho\int_c\left\{\frac{\partial\Omega_1}{\partial t}+\frac{\partial\bar{\Omega}_1}{\partial t}\right\}d\bar{z},$$

where

$$\Omega_1 = -\int_{s(z_0)}^{s(z_1)} \frac{im_s}{2\pi} \log \left\{ \frac{1-e^{-i(\zeta-\bar{\zeta}_s)}}{1-e^{i(\zeta-\bar{\zeta}_s)}} \right\} ds.$$

Writing this in the form

$$\Omega_1 = \int_{s(z_0)}^{s(z_1)} m_s f(\zeta_s, \bar{\zeta}_s) ds,$$

where

$$f(\zeta_s, \bar{\zeta}_s) = -\frac{i}{2\pi} \log \left\{ \frac{1-e^{-i(\zeta-\bar{\zeta}_s)}}{1-e^{i(\zeta-\bar{\zeta}_s)}} \right\},$$

and $\zeta_s, \bar{\zeta}_s$ are functions of s , we notice that $s(z_0)$ is the position of the trailing edge of the aerofoil, so that referred to the moving axes it has a constant value. To find $\partial\Omega_1/\partial t$, Ω_1 must be transformed to fixed axes in space by writing for the moment

$$s' = Z - s,$$

where Z is the position of the moving origin of the aerofoil referred to the fixed axes. Then

$$\Omega_1 = -\int_{Z-s(z_0)}^{Z-s(z_1)} m_s f(\zeta_s, \bar{\zeta}_s) ds',$$

where Z is now a function of the time. Then, differentiating with respect to t ,

$$\frac{\partial \Omega_1}{\partial t} = - \int_{Z-s(s_0)}^{Z-s(s_1)} \frac{\partial}{\partial t} \{m_s f(\zeta_s \bar{\zeta}_s)\} ds' - \frac{dZ}{dt} \{m_s f(\zeta_s \bar{\zeta}_s)\}_{s'=Z-s(s_1)} + \frac{dZ}{dt} \{m_s f(\zeta_s \bar{\zeta}_s)\}_{s'=Z-s(s_0)}.$$

In the old axes this means that

$$\frac{\partial \Omega_1}{\partial t} = \int_{s(s_0)}^{s(s_1)} \frac{\partial}{\partial t} \{m_s f(\zeta_s \bar{\zeta}_s)\} ds - \frac{dZ}{dt} m_1 f(\zeta_1 \bar{\zeta}_1) + \frac{dZ}{dt} m_0 f(\zeta_0 \bar{\zeta}_0),$$

where

$$m_1 = m_s(z_1), \quad \zeta_1 = \zeta(z_1), \quad \text{etc.}$$

It is usually assumed that at the end of the trail the vorticity is zero, or $m_1 = 0$, so that the second term may be dispensed with. Also assuming in the usual way that the increase in K counterbalances the circulation in the addition to the wake, then

$$\frac{dZ}{dt} m_0 = - \frac{dK}{dt}.$$

Thus

$$\frac{\partial \Omega_1}{\partial t} = \int_{s(s_0)}^{s(s_1)} \dot{m}_s f(\zeta_s \bar{\zeta}_s) ds + \int_{s(s_0)}^{s(s_1)} m_s \left\{ \frac{\partial f}{\partial \zeta_s} \dot{\zeta}_s + \frac{\partial f}{\partial \bar{\zeta}_s} \dot{\bar{\zeta}}_s \right\} ds - K f(\zeta_0 \bar{\zeta}_0),$$

and the contribution to the force is then

$$- \frac{1}{2} \rho \int_c \left(\frac{\partial \Omega_1}{\partial t} + \overline{\frac{\partial \Omega_1}{\partial t}} \right) d\bar{z}.$$

The contribution to this integral of the first two terms in the expression for Ω_1 above has been calculated in II as

$$- \rho \int_{s(s_0)}^{s(s_1)} \dot{m}_s \{z_s - c\bar{a}_1 - c\bar{a}_0 e^{i\zeta_s} + c\bar{a}_0 e^{i\zeta_s}\} ds - \rho \int_{s(s_0)}^{s(s_1)} m_s \{ \dot{\zeta}_s + i c\bar{a}_0 (\dot{\zeta}_s e^{i\zeta_s} - \dot{\bar{\zeta}}_s e^{i\bar{\zeta}_s}) \} ds.$$

The last term contributes

$$\frac{1}{2} \rho K \int_c \{f(\zeta_0 \bar{\zeta}_0) + \overline{f(\zeta_0 \bar{\zeta}_0)}\} d\bar{z},$$

and since ζ_0 lies on the cylinder, it is real and equal to $\bar{\zeta}_0$, so that

$$f(\zeta_0 \bar{\zeta}_0) = - \frac{i}{2\pi} \log \left(\frac{1 - e^{-i(\zeta - \zeta_0)}}{1 - e^{i(\zeta - \zeta_0)}} \right) = (\pi - \zeta + \zeta_0)/2\pi.$$

Thus the integral is

$$\frac{\rho K}{2\pi} \int_c (\pi - \zeta + \zeta_0) d\bar{z} = - \frac{\rho K}{2\pi} \int_{-\pi}^{+\pi} (\pi - \zeta + \zeta_0) \frac{d\bar{z}}{d\zeta} d\zeta,$$

which on integration by parts gives

$$- \rho K c \left\{ \bar{a}_0 + \sum_{n=1}^{\infty} (-)^n \bar{a}_n \right\}.$$

The complete expression for the forces on the aerofoil, using other results found in II, is therefore to be interpreted as

$$\begin{aligned}
 (3.1) \quad Y + iX = & -\pi\rho i\{Ac^2\ddot{w} - 2\bar{B}c^2\dot{w} - i\dot{\omega}c^2\bar{D}\} \\
 & -\pi\rho\omega\{Ac^2\bar{w} - 2\bar{B}c^2w - i\omega c^2\bar{D}\} \\
 & -\rho\int_{s(s_0)}^{s(s_1)} \dot{m}_s\{\bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\ell_s} + c\bar{a}_0 e^{i\ell_s}\} ds \\
 & -\rho\int_{s(s_0)}^{s(s_1)} m_s\{\dot{z}_s + i c\bar{a}_0(\dot{\ell}_s e^{i\ell_s} - \dot{\ell}_s e^{i\ell_s})\} ds \\
 & +\rho\bar{w}K - \rho i\omega c\bar{a}_0\int_{s(s_0)}^{s(s_1)} m_s\{e^{i\ell_s} - e^{i\ell_s}\} ds \\
 & +\rho\int_{s(s_0)}^{s(s_1)} m_s(\dot{z}_s + \bar{w}) ds,
 \end{aligned}$$

the two terms in K cancelling each other out, and remembering also that

$$K + \int_{s(s_0)}^{s(s_1)} m_s ds = 0.$$

4. The question now arises as to the interpretation of \dot{m}_s . Again since the axes are moving axes, it is not simply a local derivative with respect to time at a fixed distance from the origin, but a total derivative of the form

$$\dot{m}_s = \frac{\partial m_s}{\partial t} + V_s \frac{\partial m_s}{\partial s},$$

where V_s is the velocity of the vortex relative to the cylinder.

This value as it stands could be inserted in the above results and they would still be quite general, since V_s could be assumed to be a function of s . To reduce them in the special cases treated, to a form equivalent to that given by von Karman and Burgers, it must be assumed that the vortices remain at rest in the fluid once they are shed from the tail, and that therefore V_s is a constant independent of s , and in fact equal to w .^{*} This means that the last term in the expression (3.1) for $Y + iX$ is zero and that the integral

$$-\rho\int_{s(s_0)}^{s(s_1)} \dot{m}_s\{\bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\ell_s} + c\bar{a}_0 e^{i\ell_s}\} ds,$$

$$\begin{aligned}
 \text{reduces to} \quad & -\rho\int_{s(s_0)}^{s(s_1)} \frac{\partial m_s}{\partial t}\{\bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\ell_s} + c\bar{a}_0 e^{i\ell_s}\} ds \\
 & -\rho V_s\int_{s(s_0)}^{s(s_1)} \frac{\partial m_s}{\partial s}\{\bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\ell_s} + c\bar{a}_0 e^{i\ell_s}\} ds.
 \end{aligned}$$

* Some justification for such an assumption is given in the Appendix.

Integrating the second term by parts, this becomes

$$\begin{aligned} & -\rho \int_{s(z_0)}^{s(z_1)} \frac{\partial m_s}{\partial t} \{ \bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\zeta_s} + c\bar{a}_0 e^{i\zeta_s} \} ds \\ & - \rho V_s [m_s \{ \bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\zeta_s} + c\bar{a}_0 e^{i\zeta_s} \}]_{s(z_0)}^{s(z_1)} \\ & + \rho \int_{s(z_0)}^{s(z_1)} m_s \{ \dot{z}_s + i c \bar{a}_0 (\dot{\zeta}_s e^{i\zeta_s} - \dot{\bar{\zeta}}_s e^{i\bar{\zeta}_s}) \} ds, \end{aligned}$$

or, assuming as before that $m_s(z_1) = 0$ and

$$V_s m_s(z_0) = -\frac{dK}{dt},$$

the integrated term is $-\rho K \{ \bar{z}_0 - c\bar{a}_1 - c\bar{a}_0 e^{i\zeta_0} + c\bar{a}_0 e^{i\zeta_0} \},$

which when $\zeta_0 = \pi$ reduces to

$$\rho K \left\{ \bar{a}_0 + \sum_{n=2}^{\infty} (-)^n \bar{a}_n \right\}.$$

Thus finally

$$\begin{aligned} (4.1) \quad Y + iX &= -\pi \rho i \{ A c^2 \dot{w} - 2 \bar{B} c^2 \dot{w} - i \omega c^3 \bar{D} \} \\ & - \pi \rho \omega \{ A c^2 \bar{w} - 2 \bar{B} c^2 w - i \omega c^3 \bar{D} \} \\ & + \rho K c \left\{ \bar{a}_0 + \sum_{n=2}^{\infty} (-)^n \bar{a}_n \right\} \\ & - \rho \int_{s(z_0)}^{s(z_1)} \frac{\partial m_s}{\partial t} \{ \bar{z}_s - c\bar{a}_1 - c\bar{a}_0 e^{i\zeta_s} + c\bar{a}_0 e^{i\zeta_s} \} ds \\ & + \rho \bar{w} K - \rho i \omega c \bar{a}_0 \int_{s(z_0)}^{s(z_1)} m_s (e^{i\zeta_s} - e^{i\bar{\zeta}_s}) ds. \end{aligned}$$

5. In a similar way to the above, the acceleration terms in the couple due to the variations in the positions and strengths of the vortices are given by

$$-\frac{1}{2} \rho \int_c \left\{ \frac{\partial \Omega_1}{\partial t} + \overline{\frac{\partial \Omega_1}{\partial t}} \right\} z d\bar{z},$$

where

$$\Omega_1 = \int_{s(z_0)}^{s(z_1)} m_s f(\zeta_s, \bar{\zeta}_s) ds.$$

As before one finds an extra term of the form

$$\frac{1}{2} \rho K \int_c \{ f(\zeta_0, \bar{\zeta}_0) + \overline{f(\zeta_0, \bar{\zeta}_0)} \} z d\bar{z},$$

which is easily seen to reduce to

$$\frac{1}{2} \rho K c^2 \sum_{n=1}^{\infty} (-)^n (b_n + \bar{b}_n),$$

giving the expression for the couple Γ as

$$\begin{aligned} & -\pi\rho Cc^4\dot{\omega} - \frac{1}{2}\pi\rho c^2 i(D\ddot{w} - \bar{D}w) - \frac{1}{2}\pi\rho ic^2 w\bar{w}(A - \bar{A}) \\ & + \pi\rho ic^2 \{\bar{B}w^2 - B\bar{w}^2\} - \frac{1}{2}\pi\rho\omega c^2 \{\bar{D}w + D\bar{w}\} \\ & - \frac{1}{2}\rho c^2 \int_{s(s_0)}^{s(s_1)} \dot{m}_s \sum_{n=1}^{\infty} (b_n e^{n i \zeta_s} + \bar{b}_n e^{-n i \zeta_s}) ds \\ & + \frac{1}{2}\rho \int_{s(s_0)}^{s(s_1)} m_s \left[\frac{d}{dt} (z_s \bar{z}_s) - ic^2 \left\{ \xi_s \sum_{n=1}^{\infty} n b_n e^{n i \zeta_s} - \bar{\xi}_s \sum_{n=1}^{\infty} n \bar{b}_n e^{-n i \zeta_s} \right\} \right] ds \\ & + \frac{1}{2}\rho c a_0 \bar{w} \int_{s(s_0)}^{s(s_1)} m_s \{e^{-i \zeta_s} - e^{-i \bar{\zeta}_s}\} ds + \frac{1}{2}\rho c \bar{a}_0 w \int_{s(s_0)}^{s(s_1)} m_s \{e^{i \zeta_s} - e^{i \bar{\zeta}_s}\} ds. \end{aligned}$$

With the same assumptions as before, putting

$$\dot{m}_s = \frac{\partial m_s}{\partial t} + V_s \frac{\partial m_s}{\partial s},$$

the integral

$$- \frac{1}{2}\rho c^2 \int_{s(s_0)}^{s(s_1)} \dot{m}_s \sum_{n=1}^{\infty} (b_n e^{n i \zeta_s} + \bar{b}_n e^{-n i \zeta_s}) ds$$

becomes

$$\begin{aligned} & -\frac{1}{2}\rho c^2 \int_{s(s_0)}^{s(s_1)} \frac{\partial m_s}{\partial t} \sum_{n=1}^{\infty} (b_n e^{n i \zeta_s} + \bar{b}_n e^{-n i \zeta_s}) ds - \frac{1}{2}\rho c^2 K \sum_{n=1}^{\infty} (-)^n (b_n + \bar{b}_n) \\ & + \frac{1}{2}i\rho c^2 \int_{s(s_0)}^{s(s_1)} m_s \left\{ \xi_s \sum_{n=1}^{\infty} n b_n e^{n i \zeta_s} - \bar{\xi}_s \sum_{n=1}^{\infty} n \bar{b}_n e^{-n i \zeta_s} \right\} ds, \end{aligned}$$

and the couple thus reduces to

$$\begin{aligned} (5.1) \quad & -\pi\rho Cc^4\dot{\omega} - \frac{1}{2}\pi\rho c^2 i(D\ddot{w} - \bar{D}w) - \frac{1}{2}\pi\rho ic^2 w\bar{w}(A - \bar{A}) \\ & + \pi\rho ic^2 \{\bar{B}w^2 - B\bar{w}^2\} - \frac{1}{2}\pi\rho\omega c^2 \{\bar{D}w + D\bar{w}\} \\ & - \frac{1}{2}K\rho c^2 \sum_{n=1}^{\infty} (-)^n (b_n + \bar{b}_n) - \frac{1}{2}\rho c^2 \int_{s(s_0)}^{s(s_1)} \frac{\partial m_s}{\partial t} \sum_{n=1}^{\infty} (b_n e^{n i \zeta_s} + \bar{b}_n e^{-n i \zeta_s}) ds \\ & + \frac{1}{2}\rho \int_{s(s_0)}^{s(s_1)} m_s \frac{d}{dt} (z_s \bar{z}_s) + \frac{1}{2}\rho c a_0 \bar{w} \int_{s(s_0)}^{s(s_1)} m_s (e^{-i \zeta_s} - e^{-i \bar{\zeta}_s}) ds \\ & + \frac{1}{2}\rho c \bar{a}_0 w \int_{s(s_0)}^{s(s_1)} m_s (e^{i \zeta_s} - e^{i \bar{\zeta}_s}) ds. \end{aligned}$$

In the special case of the circular arc treated by von Karman & Burgers, for which

$$a_0 = 1, \quad a_1 = i\epsilon, \quad a_2 = 1, \quad a_3 = -i\epsilon \quad (n > 3, a_n = 0),$$

then, to a first approximation,

$$b_0 = 2, \quad b_1 = -i\epsilon, \quad b_2 = 1, \quad b_3 = -i\epsilon \quad (n > 3, b_n = 0),$$

and

$$A = 2, \quad B = 1, \quad C = 2, \quad D = -2i\epsilon.$$

Also $u = V \cos \alpha$, $v = -V \sin \alpha - U$, $\omega = 0$;

where α is small, and then the lift Y reduces to

$$4\pi\rho c^2 \dot{V}\alpha + 2K\rho c + \rho \int_{2c}^{\infty} \frac{\partial m_s}{\partial t} \{s - \sqrt{(s^2 - 4c^2)}\} ds + \rho VK,$$

and the couple Γ about the centre is

$$-2\pi\rho c^2 \epsilon \dot{V} + 4\pi\rho c^2 V^2 \alpha - K\rho c^2 \\ - \rho \int_{2c}^{\infty} \frac{\partial m_s}{\partial t} \left\{ \frac{1}{2}s^2 - c^2 - \frac{1}{2}s\sqrt{(s^2 - 4c^2)} \right\} + \rho V \int_{2c}^{\infty} m_s \{s - \sqrt{(s^2 - 4c^2)}\} ds.$$

These results now agree in every detail with those given by von Karman & Burgers.

THE SMALL OSCILLATION PROBLEM

6. In their paper, mentioned previously, von Karman & Sears, following Wagner, Glauert and others, discuss certain cases of small oscillation of an aerofoil about the steady forward motion, on the basis of their approach to the theory, under the impression that it has certain advantages in the direction of physical clarity over the original approach by the piecemeal method of von Karman & Burgers. In so far, however, as their method is indirect and still somewhat disconnected in character when dealing with the small oscillation about different types of steady motion, it has certain disadvantages as compared with the much more general results obtainable by my own direct method, which has the additional advantage that it enables a precise estimate to be made at each stage of the effect of the assumptions adopted, on the first form of the result.

Their method is to assume a periodic solution of the general equations of small oscillation in the complex forms

$$U + p = U + U_0 e^{i\lambda t}, \quad q = V_0 e^{i\lambda t}, \quad \omega = \omega_0 e^{i\lambda t},$$

and then to assume that the wake extends from the trailing edge straight behind the aerofoil, and that the distribution of vorticity in the wake is specified by

$$m_s = g \exp \left[\lambda \left(t - \frac{2cs}{U} \right) \right],$$

where $2cs$ is the distance along the wake from the trailing edge. Each of these small oscillations p, q, ω are dealt with separately and the coefficients U_0, V_0, ω_0 are assumed to be proportional to U . The equations of their theory then determine the lift and moment in terms of U .

I propose now to show that by my direct method this discussion may be extended considerably. In the first place the oscillations can all be assumed to be taking place simultaneously. Then without making the assumption that U_0, V_0, ω_0 are proportional to U , it is possible to discuss the dynamics of this oscillation, and in fact derive the period equation for λ corresponding to that obtained in III. By comparing these

two corresponding equations for λ , the one excluding and the other including the effect of the wake, the effect of the inclusion of the wake on the stability of the aerofoil may be discovered.

Proceeding then along these lines, the aerofoil is taken to be a flat plate moving forward at an angle of attack α . Then

$$z = 2c \cos \zeta,$$

and any point on the trail is given by

$$z_s = 2c \cos \zeta_s = -2c - 2cse^{-i\alpha}.$$

The components u , v , ω of the velocity are

$$(6.1) \quad \begin{cases} u = U \cos \alpha + U_0 e^{i\lambda} \cos \alpha - \frac{U\omega_0}{\lambda} e^{i\lambda} \sin \alpha + V_0 e^{i\lambda} \sin \alpha, \\ v = -U \sin \alpha - U_0 e^{i\lambda} \sin \alpha - \frac{U\omega_0}{\lambda} e^{i\lambda} \cos \alpha + V_0 e^{i\lambda} \cos \alpha, \\ \omega = \omega_0 e^{i\lambda}. \end{cases}$$

The Joukowski condition at the trailing edge reduces in this case to

$$-2cv + 2c^2\omega - \frac{K}{2\pi} - \int_0^\infty \frac{m_s}{2\pi} \left(\frac{e^{i\zeta_s}}{1 + e^{i\zeta_s}} + \frac{e^{-i\zeta_s}}{1 + e^{-i\zeta_s}} \right) ds = 0.$$

Now

$$z_s = c(e^{i\zeta_s} + e^{-i\zeta_s}) = -2c - 2cse^{-i\alpha}.$$

Therefore

$$e^{i\zeta_s} = -1 - se^{-i\alpha} + \sqrt{(2se^{-i\alpha} + s^2e^{-2i\alpha})},$$

$$e^{-i\zeta_s} = -1 - se^{i\alpha} - \sqrt{(2se^{i\alpha} + s^2e^{2i\alpha})},$$

$$e^{i\zeta_s} = -1 - se^{i\alpha} + \sqrt{(2se^{i\alpha} + s^2e^{2i\alpha})},$$

$$e^{-i\zeta_s} = -1 - se^{-i\alpha} - \sqrt{(2se^{-i\alpha} + s^2e^{-2i\alpha})},$$

and the Joukowski condition therefore becomes

$$-2cv + 2c^2\omega - \frac{K}{2\pi} - \int_0^\infty \frac{m_s}{2\pi} \left[\frac{-1 - se^{-i\alpha} - \sqrt{(2se^{-i\alpha} + s^2e^{-2i\alpha})}}{-se^{-i\alpha} - \sqrt{(2se^{-i\alpha} + s^2e^{-2i\alpha})}} + \frac{-1 - se^{i\alpha} - \sqrt{(2se^{i\alpha} + s^2e^{2i\alpha})}}{-se^{i\alpha} - \sqrt{(2se^{i\alpha} + s^2e^{2i\alpha})}} \right] ds = 0,$$

which can be reduced to

$$-2cv + 2c^2\omega - \frac{K}{2\pi} - \int_0^\infty \frac{m_s}{2\pi} \left(1 - \frac{\sqrt{(2se^{-i\alpha} + s^2e^{-2i\alpha})}}{2se^{-i\alpha}} - \frac{\sqrt{(2se^{i\alpha} + s^2e^{2i\alpha})}}{2se^{i\alpha}} \right) ds = 0,$$

or putting

$$s_1 = 1 + se^{-i\alpha}, \quad s_2 = 1 + se^{i\alpha}, \quad s_3 = \bar{s}_1, \quad ds_1 = e^{-i\alpha} ds, \quad ds_2 = e^{i\alpha} ds,$$

$$\text{it is simply } -2cv + 2c^2\omega - \frac{K}{2\pi} - \int_0^\infty \frac{m_s}{2\pi} \left(1 - \frac{\sqrt{(s_1^2 - 1)}}{2(s_1 - 1)} - \frac{\sqrt{(s_2^2 - 1)}}{2(s_2 - 1)} \right) ds = 0,$$

where m_s and ds must be expressed as functions of s_1 and s_2 respectively for the two different parts of the integral.

7. Considering first the integral involving s_1 , then

$$m_s = g \exp \left[\lambda \left(t - \frac{2cs}{U} \right) \right] = g \exp \left[\lambda \left(t - \frac{2c(s_1-1)e^{i\alpha}}{U} \right) \right],$$

so that

$$\begin{aligned} \int_0^\infty \frac{m_s \sqrt{(s_1^2-1)}}{2\pi 2(s_1-1)} ds &= \int_1^\infty \frac{g}{2\pi} \exp \left[\lambda \left(t + \frac{2ce^{i\alpha}}{U} \right) \right] \exp \left[-\frac{2c\lambda e^{i\alpha}}{U} s_1 \right] \frac{\sqrt{(s_1^2-1)} e^{i\alpha}}{2(s_1-1)} ds_1 \\ &= \frac{g}{4\pi} \exp \left[i\alpha + \lambda \left(t + \frac{2ce^{i\alpha}}{U} \right) \right] \int_1^\infty \exp \left[-\frac{2c\lambda e^{i\alpha}}{U} s_1 \right] \frac{(s_1+1)}{\sqrt{(s_1^2-1)}} ds_1. \end{aligned}$$

The integral in the expression on the right-hand side can be expressed as the sum of two modified Bessel functions of the second kind of argument $z = \frac{2c\lambda e^{i\alpha}}{U}$, namely,

$$K_0(z) = \int_1^\infty \frac{e^{-zs}}{\sqrt{(s^2-1)}} ds \quad \text{and} \quad K_1(z) = -K'_0(z) = \int_1^\infty \frac{se^{-zs}}{\sqrt{(s^2-1)}} ds,$$

and then

$$\int_0^\infty \frac{m_s \sqrt{(s_1^2-1)}}{2\pi 2(s_1-1)} ds = \frac{g}{4\pi} \exp \left[i\alpha + \lambda \left(t + \frac{2ce^{i\alpha}}{U} \right) \right] \left\{ K_0 \left(\frac{2c\lambda e^{i\alpha}}{U} \right) + K_1 \left(\frac{2c\lambda e^{i\alpha}}{U} \right) \right\}.$$

Similarly,

$$\int_0^\infty \frac{m_s \sqrt{(s_2^2-1)}}{2\pi 2(s_2-1)} ds = \frac{g}{4\pi} \exp \left[-i\alpha + \lambda \left(t + \frac{2ce^{-i\alpha}}{U} \right) \right] \left\{ K_0 \left(\frac{2c\lambda e^{-i\alpha}}{U} \right) + K_1 \left(\frac{2c\lambda e^{-i\alpha}}{U} \right) \right\}.$$

Writing $\frac{2c\lambda}{U} e^{i\alpha} = \mu e^{i\alpha} = \mu_+$, $\frac{2c\lambda}{U} e^{-i\alpha} = \mu e^{-i\alpha} = \mu_-$,

i.e. $K_0 \left(\frac{2c\lambda e^{i\alpha}}{U} \right) = K_0(\mu_+)$, $K_0 \left(\frac{2c\lambda e^{-i\alpha}}{U} \right) = K_0(\mu_-)$,

with a similar notation for $K_1 \left(\frac{2c\lambda e^{i\alpha}}{U} \right)$, etc., the Joukowski condition at the trailing edge becomes

$$\begin{aligned} -2cv + 2c^2\omega - \frac{K}{2\pi} - \int_0^\infty \frac{m_s ds}{2\pi} \\ + \frac{g}{4\pi} e^{i\alpha + \lambda t + \mu_+} \{K_0(\mu_+) + K_1(\mu_+)\} + \frac{g}{4\pi} e^{-i\alpha + \lambda t + \mu_-} \{K_0(\mu_-) + K_1(\mu_-)\} = 0. \end{aligned}$$

Further, suppose that when the motion is steady there is no trail and that the circulation is such that the Joukowski condition is satisfied. Let K_s be the value of the circulation in this case and let the additional circulation associated with the oscillation be K' . For steady motion in which the U_0 , V_0 , ω_0 in equations (6.1) are all zero, the Joukowski condition is

$$2cU \sin \alpha - \frac{K_s}{2\pi} = 0 \quad \text{or} \quad K_s = 4\pi cU \sin \alpha.$$

For the oscillatory motion

$$\begin{aligned}
 & 2cU \sin \alpha + 2cU_0 e^{\lambda t} \sin \alpha + \frac{2cU\omega_0 e^{\lambda t}}{\lambda} \cos \alpha - 2cV_0 e^{\lambda t} \cos \alpha \\
 & + 2c^2\omega_0 e^{\lambda t} - \frac{(K_s + K')}{2\pi} - \int_0^\infty \frac{m_s ds}{2\pi} \\
 & + \frac{g}{4\pi} e^{i\alpha + \lambda t + \mu_+} \{K_0(\mu_+) + K_1(\mu_+)\} + \frac{g}{4\pi} e^{-i\alpha + \lambda t + \mu_-} \{K_0(\mu_-) + K_1(\mu_-)\} = 0.
 \end{aligned}$$

As usual the assumption is now made that the total circulation remains the same, or that

$$K' + \int_0^\infty m_s ds = 0,$$

and substituting for K_s this gives a value for g in terms of U_0 , V_0 , ω_0 as follows:

$$\begin{aligned}
 (7.1) \quad & 2cU_0 \sin \alpha + \frac{2cU\omega_0}{\lambda} \cos \alpha - 2cV_0 \cos \alpha + 2c^2\omega_0 \\
 & + \frac{g}{4\pi} [e^{i\alpha + \lambda t + \mu_+} \{K_0(\mu_+) + K_1(\mu_+)\} + e^{-i\alpha + \lambda t + \mu_-} \{K_0(\mu_-) + K_1(\mu_-)\}] = 0.
 \end{aligned}$$

The value of K' may also be expressed in terms of g , since

$$K' = - \int_0^\infty m_s ds = - \int_0^\infty g e^{\lambda t - \mu s} ds = - \frac{g}{\mu} e^{\lambda t}.$$

Assembling the various quantities now required in the calculation of the forces and couple on the aerofoil, then

$$\begin{aligned}
 u &= U \cos \alpha + U_0 e^{\lambda t} \cos \alpha - \frac{U\omega_0}{\lambda} e^{\lambda t} \sin \alpha + V_0 e^{\lambda t} \sin \alpha, \\
 v &= -U \sin \alpha - U_0 e^{\lambda t} \sin \alpha - \frac{U\omega_0}{\lambda} e^{\lambda t} \cos \alpha + V_0 e^{\lambda t} \cos \alpha, \\
 \omega &= \omega_0 e^{\lambda t}, \\
 \dot{u} &= U_0 \lambda e^{\lambda t} \cos \alpha - U\omega_0 e^{\lambda t} \sin \alpha + V_0 \lambda e^{\lambda t} \sin \alpha, \\
 \dot{v} &= -U_0 \lambda e^{\lambda t} \sin \alpha - U\omega_0 e^{\lambda t} \cos \alpha + V_0 \lambda e^{\lambda t} \cos \alpha, \\
 \dot{\omega} &= \omega_0 \lambda e^{\lambda t}, \\
 uv &= -U^2 \sin \alpha \cos \alpha - 2UU_0 e^{\lambda t} \sin \alpha \cos \alpha - \frac{U^2\omega_0 e^{\lambda t}}{\lambda} \cos 2\alpha + UV_0 e^{\lambda t} \cos 2\alpha, \\
 u\omega &= U\omega_0 e^{\lambda t} \cos \alpha, \\
 v\omega &= -U\omega_0 e^{\lambda t} \sin \alpha, \\
 K_s &= 4\pi cU \sin \alpha, \\
 K' &= -\frac{g}{\mu} e^{\lambda t},
 \end{aligned}$$

$$m_s = ge^{\lambda t - \mu s} = ge^{\lambda t + (1-s_1)\mu_+} = ge^{\lambda t + (1-s_2)\mu_-},$$

$$\frac{\partial m_s}{\partial t} = g\lambda e^{\lambda t - \mu s} = g\lambda e^{\lambda t + (1-s_1)\mu_+} = g\lambda e^{\lambda t + (1-s_2)\mu_-},$$

$$z_s = -2c - 2cse^{-i\alpha},$$

$$\dot{z}_s = -2cse^{-i\alpha} = -Ue^{-i\alpha}, \quad \dot{\bar{z}}_s = -2cse^{i\alpha} = -Ue^{i\alpha},$$

$$e^{2i\zeta_s} = 2s_1^2 - 1 - 2s_1\sqrt{(s_1^2 - 1)}, \quad e^{-2i\bar{\zeta}_s} = 2s_2^2 - 1 - 2s_2\sqrt{(s_2^2 - 1)}.$$

8. From the general expressions (4.1) and (5.1) for the forces and couple on the aerofoil, then, for the flat plate,

$$Y = -4\pi\rho c^2\dot{v} + 2\rho c\dot{K}' + \rho u(K_s + K') - \rho R \int_{s(s_2)}^{s(s_1)} \frac{\partial m_s}{\partial t} \{z_s - \bar{a}_1 - \bar{a}_0 e^{i\zeta_s} + a_0 e^{i\bar{\zeta}_s}\} ds,$$

R as usual standing for the real part of the integral. In terms of s_1 and s_2 this reduces to

$$Y = -4\pi\rho c^2\dot{v} + 2\rho c\dot{K}' + \rho u(K_s + K') \\ + \rho c R \int_1^\infty \frac{\partial m_{s_1}}{\partial t} \{s_1 - \sqrt{(s_1^2 - 1)}\} e^{i\alpha} ds_1 + \rho c R \int_1^\infty \frac{\partial m_{s_2}}{\partial t} \{s_2 - \sqrt{(s_2^2 - 1)}\} e^{-i\alpha} ds_2,$$

and since s_2 is the conjugate of s_1 , the sum of the two integrals is purely real. On substituting the values of the various quantities and neglecting squares and products of U_0, V_0, ω_0 , then

$$Y = 4\pi\rho c^3 U_0 \lambda e^{\lambda t} \sin \alpha + 4\pi\rho c^3 U \omega_0 e^{\lambda t} \cos \alpha - 4\pi\rho c^2 V_0 \lambda e^{\lambda t} \cos \alpha \\ - \rho g U e^{\lambda t} + 4\pi\rho c U^2 \sin \alpha \cos \alpha + 4\pi\rho c U U_0 e^{\lambda t} \sin \alpha \cos \alpha \\ - \frac{4\pi\rho c U^2 \omega_0 e^{\lambda t}}{\lambda} \sin^2 \alpha + 4\pi\rho c U V_0 e^{\lambda t} \sin^2 \alpha - \frac{\rho g U e^{\lambda t}}{\mu} \cos \alpha \\ + \rho c \lambda g e^{i\alpha + \lambda t + \mu_+} \int_1^\infty e^{-s_1 \mu_+} \{s_1 - \sqrt{(s_1^2 - 1)}\} ds_1 \\ + \rho c \lambda g e^{-i\alpha + \lambda t + \mu_-} \int_1^\infty e^{-s_2 \mu_-} \{s_2 - \sqrt{(s_2^2 - 1)}\} ds_2.$$

It is necessary, therefore, to evaluate an integral of the form

$$\int_1^\infty e^{-zs} \{s - \sqrt{(s^2 - 1)}\} ds,$$

which on integration by parts reduces to

$$\frac{1}{z} e^{-z} + \frac{1}{z^2} e^{-z} - \frac{1}{z} \int_1^\infty \frac{e^{-zs}}{\sqrt{(s^2 - 1)}} ds,$$

and the last integral has been evaluated previously as $K_1(z)$, so that

$$\int_1^\infty e^{-zs} \{s - \sqrt{(s^2 - 1)}\} ds = \frac{e^{-z}}{z} + \frac{e^{-z}}{z^2} - \frac{K_1(z)}{z}.$$

Therefore the expression for Y reduces to

$$(8.1) \quad Y = 4\pi\rho c^2\lambda e^{\lambda t}(U_0 \sin \alpha - V_0 \cos \alpha) + 4\pi\rho c^2 U \omega_0 e^{\lambda t} \cos \alpha \\ + 4\pi\rho c U \sin \alpha \cos \alpha (U + U_0 e^{\lambda t}) - \frac{4\pi\rho c U^2 \omega_0 e^{\lambda t}}{\lambda} \sin^2 \alpha + 4\pi\rho c U V_0 e^{\lambda t} \sin^2 \alpha \\ - \frac{1}{2}\rho g U e^{\lambda t} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\}.$$

Similarly, the X -component of the force on the flat plate is given by

$$X = 4\pi\rho c^2 \omega v - \rho(K_s + K')v + \rho \bar{\omega} \bar{a}_0 \int_{s(z_s)}^{s(z_s)} m_s (e^{i\zeta_s} - e^{-i\zeta_s}) ds.$$

But $m_s = g e^{\lambda t - \mu s}$, and g is a function of U_0 , V_0 , ω_0 only, so that the last term can be neglected. Thus

$$(8.2) \quad X = -4\pi\rho c^2 U \omega_0 e^{\lambda t} \sin \alpha + 4\pi\rho c U (U + U_0 e^{\lambda t}) \sin^2 \alpha \\ + \frac{4\pi\rho c U^2 \omega_0 e^{\lambda t}}{\lambda} \sin \alpha \cos \alpha - 4\pi\rho c U V_0 e^{\lambda t} \sin \alpha \cos \alpha \\ - \frac{\rho g U e^{\lambda t}}{\mu} \sin \alpha.$$

9. The couple on the flat plate is, from expression (5.1),

$$\Gamma = -2\pi\rho c^4 \dot{\omega} - 4\pi\rho c^2 uv - \rho c^2 \dot{K}' - \frac{1}{2}\rho c^2 \int_0^\infty \frac{\partial m_s}{\partial t} \{e^{2i\zeta_s} + e^{-2i\zeta_s}\} ds \\ + \frac{1}{2}\rho \int_0^\infty m_s \frac{d}{dt} (z_s \bar{z}_s) ds + \frac{1}{2}\rho c (u - iv) \int_0^\infty m_s (e^{-i\zeta_s} - e^{-i\zeta_s}) ds \\ + \frac{1}{2}\rho c (u + iv) \int_0^\infty m_s (e^{i\zeta_s} - e^{i\zeta_s}) ds,$$

or substituting for z_s , \dot{z}_s , $e^{i\zeta_s}$, etc., this is

$$\Gamma = -2\pi\rho c^4 \dot{\omega} - 4\pi\rho c^2 uv - \rho c^2 \dot{K}' \\ - \frac{1}{2}\rho c^2 e^{i\alpha} \int_1^\infty \frac{\partial m_{s_1}}{\partial t} \{2s_1^2 - 1 - 2s_1 \sqrt{(s_1^2 - 1)}\} ds_1 \\ - \frac{1}{2}\rho c^2 e^{-i\alpha} \int_1^\infty \frac{\partial m_{s_2}}{\partial t} \{2s_2^2 - 1 - 2s_2 \sqrt{(s_2^2 - 1)}\} ds_2 \\ + \rho c U e^{2i\alpha} \int_1^\infty m_{s_1} s_1 ds_1 + \rho c U e^{-2i\alpha} \int_1^\infty m_{s_2} s_2 ds_2 \\ + \rho c i v e^{i\alpha} \int_1^\infty m_{s_1} s_1 ds_1 - \rho c i v e^{-i\alpha} \int_1^\infty m_{s_2} s_2 ds_2 \\ - \rho c u e^{i\alpha} \int_1^\infty m_{s_1} \sqrt{(s_1^2 - 1)} ds_1 - \rho c u e^{-i\alpha} \int_1^\infty m_{s_2} \sqrt{(s_2^2 - 1)} ds_2,$$

or with the appropriate values of m_s

$$\begin{aligned} \Gamma = & -2\pi\rho c^4\dot{\omega} - 4\pi\rho c^2 uv - \rho c^2 \dot{K}' \\ & - \rho c g e^{i\alpha + \lambda t + \mu_+} \int_1^\infty e^{-s_1\mu_+} [c\lambda\{s_1^2 - \frac{1}{2} - s_1\sqrt{(s_1^2 - 1)}\} - Ue^{i\alpha}s_1 - ivs_1 + u\sqrt{(s_1^2 - 1)}] ds_1 \\ & - \rho c g e^{-i\alpha + \lambda t + \mu_-} \int_1^\infty e^{-s_2\mu_-} [c\lambda\{s_2^2 - \frac{1}{2} - s_2\sqrt{(s_2^2 - 1)}\} - Ue^{-i\alpha}s_2 + ivs_2 + u\sqrt{(s_2^2 - 1)}] ds_2. \end{aligned}$$

An integral of the form $\int_1^\infty e^{-ss}\{s^2 - \frac{1}{2} - s\sqrt{(s^2 - 1)}\} ds$

has not yet been evaluated. Integrating by parts, then

$$\left| -\frac{e^{-ss}}{z}\{s^2 - \frac{1}{2} - s\sqrt{(s^2 - 1)}\} \right|_1^\infty + \int_1^\infty \frac{e^{-ss}}{z} \left\{ 2s - \sqrt{(s^2 - 1)} - \frac{s^2}{\sqrt{(s^2 - 1)}} \right\} ds,$$

and each part of the integral in terms of K_0 and K_1 has already been found, so that inserting these values one finally obtains

$$\int_1^\infty e^{-ss}\{s^2 - \frac{1}{2} - s\sqrt{(s^2 - 1)}\} ds = \frac{e^{-z}}{2z} + \frac{2e^{-z}}{z^2} + \frac{2e^{-z}}{z^3} - \frac{2K_1}{z^2} - \frac{K_0}{z}.$$

Inserting also the values of u , v , ω the expression for Γ becomes

$$\begin{aligned} \Gamma = & -2\pi\rho c^4\omega_0\lambda e^{\lambda t} + 4\pi\rho c^2 U(U + 2U_0 e^{\lambda t}) \sin\alpha \cos\alpha \\ & + \frac{4\pi\rho c^2 U^2\omega_0 e^{\lambda t}}{\lambda} \cos 2\alpha - 4\pi\rho c^2 UV_0 e^{\lambda t} \cos 2\alpha \\ & + \frac{2\rho c g e^{\lambda t} \sin^2\alpha}{\mu^2} \\ & - \frac{i\rho c g U e^{\lambda t}}{\mu} \sin\alpha \{e^{\mu_+} K_1(\mu_+) - e^{\mu_-} K_1(\mu_-)\} \\ & + \frac{1}{2} \rho c g U e^{\lambda t} \{e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)\}. \end{aligned}$$

10. Having found these general expressions for the forces and couple on an aerofoil moving in the manner described by equations (6.1), it would be as well perhaps to show that for the simpler cases treated by von Karman & Sears, they reduce to results which are consistent with theirs.

In the first place, in their analysis, the angle of attack α is zero, and the length of the chord of the aerofoil $2c$ is taken to be unity. This means that

$$K_0(\mu_+) = K_0(\mu_-) = K_0\left(\frac{\lambda}{U}\right), \quad K_1(\mu_+) = K_1(\mu_-) = K_1\left(\frac{\lambda}{U}\right).$$

Then considering their case I, dealing with translatory motion, put

$$U_0 = 0, \quad V_0 = -AU, \quad \omega_0 = 0,$$

whence from equation (7.1) it is found that g is given by

$$g = -2\pi A_0 U e^{\lambda U} / (K_0 + K_1).$$

This is equivalent to the value of g given on p. 385 of their paper when equation (25) is combined with the value of G_0 given for case I, remembering that their ζ , ν are equivalent to my $1 + s$, λ respectively.

Also with these particular values, it is found from equations (8.1) and (8.2) that

$$Y = \pi \rho e^{\lambda U} A_0 U \lambda - \rho g U e^{\lambda U} \{e^{\lambda U} K_1\},$$

which with the above value of g becomes

$$Y = 2\pi \rho U^2 e^{\lambda U} A_0 \left\{ \frac{K_0(\lambda/U)}{K_0(\lambda/U) + K_1(\lambda/U)} + \frac{\lambda}{2U} \right\}.$$

Also

$$X = 0$$

and
$$\Gamma = \pi \rho A_0 U^2 e^{\lambda U} + \frac{1}{2} \rho g U e^{\lambda U} K_0 = -\pi \rho A_0 U^2 e^{\lambda U} \frac{K_1(\lambda/U)}{K_0(\lambda/U) + K_1(\lambda/U)},$$

which are consistent with equations (29) and (30) on p. 385.

Similarly, the results of case II for rotational oscillation are found by putting

$$U_0 = 0, \quad V_0 = 0, \quad \omega_0 = 2UA_1,$$

giving
$$Y = 2\pi \rho U^2 A_1 e^{\lambda U} \frac{K_1}{K_0 + K_1} + 2\pi \rho U^2 A_1 e^{\lambda U} + \frac{4\pi \rho U^3 A_1 e^{\lambda U}}{\lambda} \frac{K_1}{K_0 + K_1},$$

$$X = 0,$$

and
$$\Gamma = -\frac{1}{2} \pi \rho U A_1 \lambda e^{\lambda U} + \frac{2\pi \rho U^3 A_1 e^{\lambda U}}{\lambda} - \pi \rho U^2 A_1 e^{\lambda U} \left(1 + \frac{2U}{\lambda} \right) \frac{K_0}{K_0 + K_1}.$$

These agree with those found by von Karman & Sears if the terms suggested in the footnote on p. 385 of the paper are included. The results above could, of course, have been found almost immediately by direct substitution of these special values of U_0 , V_0 , ω_0 in the original equations for the forces and couple.

STABILITY OF THE OSCILLATION

11. Returning now to the discussion of the general equations of small oscillation and their bearing on the question of stability, the general values for the force and couple components X , Y , Γ are substituted in the usual equations of motion

$$M(\ddot{u} - \omega v - y_G \dot{\omega} - x_G \omega^2) = X + X_1,$$

$$M(\ddot{v} + \omega u + x_G \dot{\omega} - y_G \omega^2) = Y + Y_1,$$

$$I_G \ddot{\omega} = -I_G n^2 \theta + \Gamma + X y_G - Y x_G.$$

These yield the usual kind of equations, two for the linear and one for the angular motion:

$$\begin{aligned} & \left\{ M\lambda \cos \alpha - 4\pi\rho c U \sin^2 \alpha - \frac{4\pi\rho U^2 \sin^2 \alpha}{G\lambda} \right\} U_0 e^{\lambda t} \\ & + \left\{ M\lambda \sin \alpha + 4\pi\rho c U \sin \alpha \cos \alpha + \frac{4\pi\rho U^2 \sin \alpha \cos \alpha}{G\lambda} \right\} V_0 e^{\lambda t} \\ & - \left\{ M y_G \lambda - 4\pi\rho c^2 U \sin \alpha + \frac{4\pi\rho c U^2 \sin \alpha \cos \alpha}{\lambda} + \frac{4\pi\rho U^2 \sin \alpha}{G\lambda} \left(\frac{U \cos \alpha}{\lambda} + c \right) \right\} \omega_0 e^{\lambda t} \\ & - 4\pi\rho c U^2 \sin^2 \alpha - X_1 = 0, \end{aligned}$$

$$\begin{aligned} & \left[M\lambda \sin \alpha + 4\pi\rho c^2 \lambda \sin \alpha + 4\pi\rho c U \sin \alpha \cos \alpha \right. \\ & \quad \left. + \frac{4\pi\rho c U}{G} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} \sin \alpha \right] U_0 e^{\lambda t} \\ & - \left[M\lambda \cos \alpha + 4\pi\rho c^2 \lambda \cos \alpha - 4\pi\rho c U \sin^2 \alpha \right. \\ & \quad \left. + \frac{4\pi\rho c U}{G} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} \cos \alpha \right] V_0 e^{\lambda t} \\ & - \left[M x_G \lambda - 4\pi\rho c^2 U \cos \alpha + 4\pi\rho c U^2 \sin^2 \alpha \right. \\ & \quad \left. - \frac{4\pi\rho c U}{G} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} \left(\frac{U \cos \alpha}{\lambda} + c \right) \right] \omega_0 e^{\lambda t} \\ & + 4\pi\rho c U^2 \sin \alpha \cos \alpha + Y_1 = 0, \end{aligned}$$

and finally

$$\begin{aligned} & \left[4\pi\rho c^2 x_G \lambda \sin \alpha - 4\pi\rho c^2 U \sin 2\alpha + 2\pi\rho c x_G U \sin 2\alpha - 4\pi\rho c y_G U \sin^2 \alpha \right. \\ & \quad - \frac{4\pi\rho c^2 U \sin \alpha}{G} \{e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)\} + \frac{4\pi\rho c x_G U \sin \alpha}{G} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} \\ & \quad - \frac{4\pi\rho c U^2 \sin^2 \alpha}{G\lambda} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} - \frac{4\pi\rho y_G U^2 \sin^2 \alpha}{G\lambda} + \frac{4\pi\rho U^2 \sin^2 \alpha}{G\lambda^2} \left. \right] U_0 e^{\lambda t} \\ & - \left[4\pi\rho c^2 x_G \lambda \cos \alpha - 4\pi\rho c^2 U \cos 2\alpha - 4\pi\rho c x_G U \sin^2 \alpha - 2\pi\rho c y_G U \sin 2\alpha \right. \\ & \quad + \frac{4\pi\rho c^2 U \cos \alpha}{G} \{e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)\} \\ & \quad + \frac{4\pi\rho c x_G U \cos \alpha}{G} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} - \frac{2\pi\rho y_G U^2 \sin 2\alpha}{G\lambda} \\ & \quad - \frac{2\pi\rho c U^2 \sin 2\alpha}{G\lambda} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} + \frac{4\pi\rho U^2 \sin^2 \alpha \cos \alpha}{G\lambda^2} \left. \right] V_0 e^{\lambda t} \end{aligned}$$

$$\begin{aligned}
& + \left[(I_G + 2\pi\rho c^4) \lambda + 4\pi\rho c^2 x_G U \cos \alpha + 4\pi\rho c^2 y_G U \sin \alpha + \frac{I_G n^2}{\lambda} \right. \\
& \quad - \frac{4\pi\rho c^2 U^2 \cos 2\alpha}{\lambda} - \frac{4\pi\rho c x_G U^2 \sin^2 \alpha}{\lambda} - \frac{2\pi\rho c y_G U^2 \sin 2\alpha}{\lambda} \\
& \quad + \frac{4\pi\rho c}{G} \left(\frac{U \cos \alpha}{\lambda} + c \right) \left\{ \frac{U^2 \sin^2 \alpha}{c\lambda^2} - \frac{iU^2 \sin \alpha}{\lambda} (e^{\mu_+} K_1(\mu_+) - e^{\mu_-} K_1(\mu_-)) \right. \\
& \quad + cU(e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)) + x_G U(e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)) - \frac{y_G U^2 \sin \alpha}{c\lambda} \left. \right\} \Big] \omega_0 e^{\lambda t} \\
& \quad - 4\pi\rho c(c - x_G) U^2 \sin \alpha \cos \alpha - 4\pi\rho c y_G U^2 \sin^2 \alpha = 0,
\end{aligned}$$

where $G = e^{i\alpha+\mu_+} \{K_0(\mu_+) + K_1(\mu_+)\} + e^{-i\alpha+\mu_-} \{K_0(\mu_-) + K_1(\mu_-)\}.$

On examination it will be found that these equations bear a close resemblance to the equations for the oscillation of the aerofoil when the effect of the wake is neglected. What has really happened on the introduction of the effect of the wake is that some of the terms have been split up, the one part of such terms remaining as it is, being that which has arisen from the circulation, and the other part now showing the effect of the wake. As can be seen, all such terms are small in comparison with the other terms involving M and I_G in the equations, and it can safely be said that the same approximations can again be taken as were taken before. This means that the equation for λ reduces to

$$\begin{aligned}
& \left\{ (I_G + 2\pi\rho c^4) \lambda + 4\pi\rho c^2 x_G U \cos \alpha + 4\pi\rho c^2 y_G U \sin \alpha \right. \\
& \quad + \frac{I_G n^2}{\lambda} - \frac{4\pi\rho c^2 U^2 \cos 2\alpha}{\lambda} - \frac{4\pi\rho c x_G U^2 \sin^2 \alpha}{\lambda} - \frac{4\pi\rho c y_G U^2 \sin 2\alpha}{\lambda} \left. \right\} \\
& \quad \times [e^{i\alpha+\mu_+} \{K_0(\mu_+) + K_1(\mu_+)\} + e^{-i\alpha+\mu_-} \{K_0(\mu_-) + K_1(\mu_-)\}] \\
& \quad + \frac{4\pi\rho U^4 \sin^2 \alpha \cos \alpha}{\lambda^3} - \frac{4\pi\rho c i U^2 \sin \alpha \cos \alpha}{\lambda^2} \{e^{\mu_+} K_1(\mu_+) - e^{\mu_-} K_1(\mu_-)\} \\
& \quad + \frac{4\pi\rho c^2 U^2 \cos \alpha}{\lambda} \{e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)\} \\
& \quad + \frac{4\pi\rho c x_G U^2 \cos \alpha}{\lambda} \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} \\
& \quad - \frac{4\pi\rho c y_G U^2 \cos \alpha \sin \alpha}{\lambda^2} + \frac{4\pi\rho c U^2 \sin^2 \alpha}{\lambda^2} - \frac{4\pi\rho c y_G U^2 \sin \alpha}{\lambda} \\
& \quad - \frac{4\pi\rho c^2 i U^2 \sin \alpha}{\lambda} \{e^{\mu_+} K_1(\mu_+) - e^{\mu_-} K_1(\mu_-)\} \\
& \quad + 4\pi\rho c^3 U \{e^{\mu_+} K_0(\mu_+) + e^{\mu_-} K_0(\mu_-)\} + 4\pi\rho c^2 x_G U \{e^{\mu_+} K_1(\mu_+) + e^{\mu_-} K_1(\mu_-)\} = 0,
\end{aligned}$$

where, of course,

$$\mu_+ = \mu e^{i\alpha} = \frac{2c\lambda}{U} e^{i\alpha}, \quad \mu_- = \mu e^{-i\alpha} = \frac{2c\lambda}{U} e^{-i\alpha}.$$

One must therefore now try to reduce this equation further to see if a condition for stability can be found. In the first place notice that in all practical problems $2c\lambda/U$ would be very small. In the modified Bessel functions K_0 and K_1 , as well as in the exponential terms $e^{\mu\pm}$, etc., the terms of lowest degree in their series form can be used. Thus, γ denoting Euler's constant 0.57, these are

$$K_0(z) = -\gamma, \quad K_1(z) = \frac{1}{z}, \quad e^{\mu\pm} = \left(1 + \frac{2c\lambda}{U} e^{i\alpha}\right).$$

Substituting these values in the coefficients required in the above equation, then

$$e^{i\alpha+\mu_+}\{K_0(\mu_+) + K_1(\mu_+)\} + e^{-i\alpha+\mu_-}\{K_0(\mu_-) + K_1(\mu_-)\} = \frac{U}{c\lambda},$$

$$e^{\mu_+}K_0(\mu_+) + e^{\mu_-}K_0(\mu_-) = -2\gamma,$$

$$e^{\mu_+}K_1(\mu_+) + e^{\mu_-}K_1(\mu_-) = \frac{U}{c\lambda} \cos \alpha + 2,$$

$$e^{\mu_+}K_1(\mu_+) - e^{\mu_-}K_1(\mu_-) = -\frac{iU}{c\lambda} \sin \alpha,$$

and the equation for λ then reduces to

$$\begin{aligned} \frac{U}{c\lambda} & \left\{ (I_G + 2\pi\rho c^4) \lambda + 4\pi\rho c^2 x_G U \cos \alpha + 4\pi\rho c^2 y_G U \sin \alpha + \frac{I_G n^2}{\lambda} \right. \\ & - \frac{4\pi\rho c^2 U^2 \cos 2\alpha}{\lambda} - \frac{4\pi\rho c x_G U^2 \sin^2 \alpha}{\lambda} - \frac{2\pi\rho c y_G U^2 \sin 2\alpha}{\lambda} \Big\} \\ & - 8\pi\rho c^2 \gamma \frac{U^2 \cos \alpha}{\lambda} + \frac{4\pi\rho x_G U^2 \cos^2 \alpha}{\lambda^2} - \frac{4\pi\rho y_G U^2 \cos \alpha \sin \alpha}{\lambda^2} \\ & - 8\pi\rho c^2 \gamma U + \frac{4\pi\rho c x_G U^2 \cos \alpha}{\lambda} + 8\pi\rho c^2 x_G U - \frac{4\pi\rho c y_G U^2 \sin \alpha}{\lambda} = 0, \end{aligned}$$

$$\begin{aligned} \text{or} \quad & (I_G + 2\pi\rho c^4 - 8\pi\rho\gamma c^4 + 8\pi\rho c^2 x_G) \lambda^2 + (16\pi\rho c^2 x_G - 8\pi\rho c^2 \gamma) U \cos \alpha \lambda \\ & + I_G n^2 - 4\pi\rho c(c - x_G) U^2 \cos 2\alpha - 4\pi\rho c y_G U^2 \sin 2\alpha = 0. \end{aligned}$$

By expressing the condition that the values of λ given by this equation must be complex if the motion is to be stable, a criterion for stability which is very similar to the one in III is obtained, where in fact the equation for λ was

$$\begin{aligned} & (I_G + 6\pi\rho c^4 + 8\pi\rho c^2 x_G) \lambda^2 + 4\pi\rho c^2(c + 4x_G) U \cos \alpha \cdot \lambda \\ & + I_G n^2 - 4\pi\rho c(c - x_G) U^2 \cos 2\alpha - 4\pi\rho c y_G U^2 \sin 2\alpha = 0. \end{aligned}$$

Of course in writing down the criterion for stability it is the constant term in the equation which is the most important, and it is seen that this is the same in both cases. If, however, the coefficients of λ and λ^2 are compared, one finds that the inclusion of the effect of the wake can be shown to influence the stability.

In fact, if x'_G is taken to be the position of the centre of gravity when including the effect of the wake, then by equating the coefficients of λ which are next in importance

$$4\pi\rho c^2(c + 4x_G) = 16\pi\rho c^2 x'_G - 8\pi\rho c^2 \gamma,$$

$$x'_G = x_G + \frac{1}{4}c(1 + 2\gamma),$$

and similarly from the coefficients of λ^2

$$x''_G = x_G + \frac{1}{2}c(1 + 2\gamma).$$

The second term is mainly important in connexion with the damping of the oscillations under review. Here it is seen that the effect of the trail is equivalent—as compared with the result without the trail—to a shift in the centre of gravity of the wing structure backwards towards the trailing edge and by a distance

$$c(1 + 2\gamma)/4 = \frac{1}{2}c \text{ approx.},$$

or one-eighth of the chord length. This is probably the most important effect of all, as it shows that unless the centre of gravity of the structure is well forward the effect of the trail might easily convert a natural damping factor (negative exponential) into an increasing factor (positive exponential) indicating instability. For stability it is necessary in fact that

$$x_G > \frac{1}{2}\gamma c > 0.28c = 0.07 \text{ (chord length)}.$$

In the third coefficient—the first in the equation—the effect is on the hydrodynamical inertia effect, itself only a first order correction to the rotational inertia of the aeroplane structure. And here again the effect is equivalent to a shift of the centre of gravity of the structure backwards towards the trailing edge and by a distance equal approximately to c or one-quarter of the chord length.

APPENDIX

On the motion of a vortex behind a moving cylinder

12. It was pointed out in the results of paragraph 4 that in order that the results there should agree with those of von Karman & Burgers, it is necessary to assume that the vortices, once they are shed from the cylinder, remain at rest in the fluid. Although the theorem relating to the motion of vortices with the fluid, given by Kelvin and Helmholtz, is well known, it seems that there has been little, if any, theoretical basis for the assumption that in the neighbourhood of the vortex situated behind a moving cylinder the velocity of the fluid is very small. I propose, therefore, to examine this assumption by finding the relative motion of a single vortex behind a plane aerofoil which is moving forward with a velocity U , at a finite angle of attack α , thereby determining the degree of approximation involved in assuming the vortex at rest. The vortex is taken to be of strength m , situated at any point in the stream behind the aerofoil.

The plane aerofoil being defined by the transformation

$$z = 2c \cos \zeta, \quad \eta = 0,$$

any point in the surrounding fluid can be taken to be on a confocal ellipse $\eta = \eta_s$, so that the position of the vortex is given by

$$z_s = 2c \cos \zeta_s.$$

The complex potential of the above motion for this particular case is then derived from previous results in I and II to be

$$\Omega = 2ciU \sin \alpha e^{i\zeta} - \frac{K\zeta}{2\pi} - \frac{im}{2\pi} \log \left(\frac{1 - e^{-i(\zeta - \zeta_s)}}{1 - e^{i(\zeta - \zeta_s)}} \right).$$

With my usual notation it is now known that $\dot{z}_s - i\omega z_s + \bar{w}$ is the conjugate complex velocity in the neighbourhood of the vortex, excluding the large motion due to the vortex itself. That is,

$$\dot{z}_s - i\omega \bar{z}_s + \bar{w} = \left(\frac{\partial \Omega}{\partial z} - \frac{\partial \Omega_0}{\partial z} \right)_{z=z_s},$$

where Ω_0 is the complex potential function of the vortex alone given by

$$\Omega_0 = -\frac{im}{2\pi} \log (z - z_s).$$

As part of the right-hand side of the above equation, one requires the limit of the derivative with respect to z of*

$$-\frac{im}{2\pi} \log \left(\frac{1 - e^{-i(\zeta - \zeta_s)}}{1 - e^{i(\zeta - \zeta_s)}} \right) + \frac{im}{2\pi} \log (z - z_s),$$

when $z = z_s$. This is found by putting $\zeta = \zeta_s + t$, where t is small and writing the derivative of the above expression in the form

$$\frac{m}{2\pi i t z'_s} \left\{ 1 - it \left(1 - \frac{1}{2} \coth \eta_s \right) - \frac{z''_s}{z'_s} t \right\} - \frac{m}{2\pi i t z'_s} \left(1 - \frac{z''_s}{z'_s} t \right),$$

or

$$\frac{m}{4\pi z'_s} \left\{ \frac{iz''_s}{z'_s} - 2 + \coth \eta_s \right\},$$

where dashes denote differentiation with respect to ζ so that

$$z'_s = (dz/d\zeta)_{\zeta=\zeta_s} \quad \text{and} \quad z''_s = (d^2\zeta/d\zeta^2)_{\zeta=\zeta_s}.$$

With $w = Ue^{i\alpha}$, $\omega = 0$, the above equation then becomes

$$(12.1) \quad \dot{z}_s + Ue^{i\alpha} = -\frac{2cU \sin \alpha e^{i\zeta_s}}{z'_s} - \frac{K}{2\pi z'_s} + \frac{m}{4\pi z'_s} \left\{ \frac{iz''_s}{z'_s} - 2 + \coth \eta_s \right\}.$$

* This is slightly different from the calculations in II. The alteration has been made after Dr Green kindly pointed out to me the error in that part of the work.

In all the previous work dealing with vortices and a trail behind an aerofoil the Joukowski condition of finite velocity at the trailing edge is used, and therefore for these results to have any relevance to this work this condition must be used. In this particular case it reduces to

$$2cU \sin \alpha - \frac{K}{2\pi} - \frac{m}{2\pi} \left\{ 1 - \frac{\sinh \eta_s}{\cosh \eta_s + \cos \xi_s} \right\} = 0.$$

It must also be noticed that before the vortex could be shed from the aerofoil there must have been at some time a change in velocity and therefore also in circulation. Suppose that before the vortex is shed the velocity and circulation are U_0 and K_0 respectively, and that the Joukowski condition at the trailing edge is satisfied, so that

$$2cU_0 \sin \alpha - \frac{K_0}{2\pi} = 0.$$

Now suppose that the velocity changes to $U = U_0 + U'$ and the circulation to $K = K_0 + K'$, and that during the change the vortex m is shed. Then since the total circulation must remain constant

$$K' + m = 0.$$

If, after the change the velocity remains constant, the Joukowski condition will be

$$2c(U_0 + U') \sin \alpha - \frac{K_0 + K'}{2\pi} - \frac{m}{2\pi} \left\{ 1 - \frac{\sinh \eta_s}{\cosh \eta_s + \cos \xi_s} \right\} = 0,$$

and, using the two conditions above,

$$\frac{m}{2\pi} = - \frac{2cU' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{\sinh \eta_s},$$

and thence

$$\frac{K'}{2\pi} = \frac{2cU' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{\sinh \eta_s}.$$

Substituting the above values for m and K in the equation (12.1), one obtains for the relative velocity of the vortex

$$\begin{aligned} \dot{z}_s = & -(U_0 + U') e^{i\alpha} - \frac{2c(U_0 + U') e^{i\alpha} \sin \alpha}{z'_s} - \frac{2cU_0 \sin \alpha}{z'_s} \\ & - \frac{cU' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{z'_s \sinh \eta_s} \left\{ \frac{iz''_s}{z'_s} + \coth \eta_s \right\}. \end{aligned}$$

From this expression it is easily seen that when the vortex is immediately behind the aerofoil, i.e. $\xi_s = \pi$, $\eta_s = 0$, the relative velocity remains finite, the infinite terms cancelling each other.

Consider now a vortex situated on the line extending from the trailing edge of the aerofoil straight behind in the negative direction of motion, so that

$$z_s = -2c - 2cse^{-i\alpha},$$

where $2cs$ is the distance from the trailing edge. Then

$$\dot{z}_s = -2cse^{i\alpha} - 2csi e^{i\alpha} \dot{\alpha} \quad \text{or} \quad \dot{z}_s e^{-i\alpha} = -2cs - 2csi \dot{\alpha}.$$

The velocity along the line is therefore the real part of $-\dot{z}_s e^{-i\alpha}$, and the velocity perpendicular to the line is the imaginary part of this expression. But from above

$$(12.2) \quad -\dot{z}_s e^{-i\alpha} = (U_0 + U') + \frac{2c\bar{z}'_s(U_0 + U') e^{i\xi_s - i\alpha} \sin \alpha}{z'_s \bar{z}'_s} + \frac{2e\bar{z}'_s U_0 e^{-i\alpha} \sin \alpha}{z'_s \bar{z}'_s} \\ + \frac{c\bar{z}'_s e^{-i\alpha} \sin \alpha (\cosh \eta_s + \cos \xi_s)}{z'_s \bar{z}'_s \sinh \eta_s} \left\{ \frac{iz'_s \bar{z}'_s}{z'_s \bar{z}'_s} + \coth \eta_s \right\},$$

where the denominators have been made all real. The following values are then to be substituted in the expression:

$$(12.3) \quad \begin{cases} z_s = 2c\{\cos \xi_s \cosh \eta_s - i \sin \xi_s \sinh \eta_s\}, \\ \bar{z}'_s = -2c\{\sin \xi_s \cosh \eta_s - i \cos \xi_s \sinh \eta_s\}, \\ z'_s \bar{z}'_s = 4c^2\{\cosh^2 \eta_s - \cos^2 \xi_s\}, \\ z'_s \bar{z}'_s = -4c^2\{\sin \xi_s \cos \xi_s - \sinh \eta_s \cosh \eta_s\}, \end{cases}$$

where ξ_s, η_s are given in terms of s by

$$\begin{cases} \cos(\xi_s + i\eta_s) = -1 - se^{-i\alpha}, \\ \text{or} \quad \begin{cases} \cos \xi_s \cosh \eta_s = -1 - s \cos \alpha \\ \sin \xi_s \sinh \eta_s = -s \sin \alpha \end{cases} \end{cases}$$

The relative velocity of the vortex perpendicular to the line of motion, being the imaginary part of the right-hand side of the above equation, is

$$\begin{aligned} & -\frac{4c^2 U_0 \sin \alpha}{z'_s \bar{z}'_s} [\sin^2 \xi_s \cosh^2 \eta_s \cos \alpha + \cos^2 \xi_s \sinh^2 \eta_s \cos \alpha - \sinh \eta_s \cosh \eta_s \cos \alpha \\ & \quad - \sin \xi_s \cos \xi_s \sin \alpha - \cos \xi_s \sinh \eta_s \cos \alpha - \sin \xi_s \cosh \eta_s \sin \alpha] \\ & -\frac{4c^2 U' \sin \alpha}{z'_s \bar{z}'_s} (\cosh \eta_s - \sinh \eta_s) \{\sin \xi_s \cosh \eta_s \sin(\xi_s - \alpha) - \cos \xi_s \sinh \eta_s \cos(\xi_s - \alpha)\} \\ & +\frac{2c^2 U' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{z'_s \bar{z}'_s \sinh^2 \eta_s} \{\cosh \eta_s (\sin \xi_s \cosh \eta_s \sin \alpha + \cos \xi_s \sinh \eta_s \cos \alpha) \\ & \quad + \sinh \eta_s (\cos \xi_s \cosh \eta_s \cos \alpha + \sin \xi_s \sinh \eta_s \sin \alpha)\} \\ & +\frac{8c^4 U' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{z'^2_s \bar{z}'^2_s \sinh \eta_s} \{2 \cos^2 \xi_s (\cos \xi_s \cosh \eta_s \cos \alpha + \sin \xi_s \sinh \eta_s \sin \alpha) \\ & \quad - 2 \cos \xi_s \cosh \eta_s \cos \alpha\}. \end{aligned}$$

By the substitution of s and α in place of ξ_s and η_s in some of the terms, this can be further reduced to

$$\begin{aligned}
 & -U_0 \sin \alpha \cos \alpha + \frac{U_0 \sin \alpha (\cos \alpha \sinh \eta_s + \sin \alpha \sin \xi_s)}{\cosh \eta_s - \cos \xi_s} \\
 & - \frac{U' \sin \alpha \cos \alpha (\cosh \eta_s - \sinh \eta_s)}{\cosh \eta_s - \cos \xi_s} + \frac{U' \sin^2 \alpha \sin \xi_s}{2 \sinh^2 \eta_s (\cosh \eta_s - \cos \xi_s)} \\
 & - \frac{U' \sin \alpha (\cosh \eta_s - \sinh \eta_s)}{(\cosh^2 \eta_s - \cos^2 \xi_s)} \{s \cos \xi_s + \sin \xi_s \sin \alpha - \sinh \eta_s \cos \alpha\} \\
 & - \frac{U' \sin \alpha (s + \cos \alpha)}{\sinh \eta_s (\cosh \eta_s - \cos \xi_s)} + \frac{U' \sin \alpha (\sin^2 \xi_s \cos \alpha + s \cos^2 \alpha - s \cos^2 \xi_s)}{\sinh \eta_s (\cosh \eta_s - \cos \xi_s) (\cosh^2 \eta_s - \cos^2 \xi_s)}.
 \end{aligned}$$

Similarly the relative velocity of the vortex along the line of motion, being the real part of $-\dot{z}_s e^{-i\alpha}$, is

$$\begin{aligned}
 (U_0 + U') - \frac{4c^2 U_0 \sin \alpha}{z'_s \bar{z}'_s} \{ \cosh^2 \eta_s \sin \alpha - \cos^2 \xi_s \sin \alpha + \sin \xi_s \cos \xi_s \cos \alpha \\
 - \sinh \eta_s \cosh \eta_s \sin \alpha + \sin \xi_s \cosh \eta_s \cos \alpha - \cos \xi_s \sinh \eta_s \sin \alpha \} \\
 - \frac{4c^2 U' \sin \alpha (\cosh \eta_s - \sinh \eta_s)}{z'_s \bar{z}'_s} \{ \sin \xi_s \cosh \eta_s \cos (\xi_s - \alpha) + \cos \xi_s \sinh \eta_s \sin (\xi_s - \alpha) \} \\
 - \frac{2c^2 U' \sin \alpha \cosh \eta_s (\cosh \eta_s + \cos \xi_s)}{z'_s \bar{z}'_s \sinh^2 \eta_s} \{ \sin \xi_s \cosh \eta_s \cos \alpha - \cos \xi_s \sinh \eta_s \sin \alpha \} \\
 + \frac{8c^4 U' \sin \alpha (\cosh \eta_s + \cos \xi_s)}{z'^2_s \bar{z}'^2_s \sinh \eta_s} \{ (\cosh^2 \eta_s + \cos^2 \xi_s) (\sin \xi_s \sinh \eta_s \cos \alpha \\
 + \cos \xi_s \sinh \eta_s \sin \alpha) - 2 \cos \xi_s \cosh \eta_s \sin \alpha \}.
 \end{aligned}$$

By using the equations (12.3) in some of the terms, this can again be reduced to

$$\begin{aligned}
 (U_0 + U') - U_0 \sin^2 \alpha - \frac{U_0 \sin \alpha (\sin \xi_s \cos \alpha - \sinh \eta_s \sin \alpha)}{(\cosh \eta_s + \cos \xi_s)} \\
 - \frac{U' \sin^2 \alpha (\cosh \eta_s - \sinh \eta_s)}{\cosh \eta_s - \cos \xi_s} \\
 + \frac{U' \sin \alpha (\cosh \eta_s - \sinh \eta_s)}{\cosh^2 \eta_s - \cos^2 \xi_s} \{s \sin \xi_s + \cos \alpha \sin \xi_s + \sinh \eta_s \sin \alpha\} \\
 + \frac{U' \sin \alpha \cos \alpha \sin \xi_s (\cosh^2 \eta_s - \sinh^2 \eta_s)}{2 \sinh^2 \eta_s (\cosh \eta_s - \cos \xi_s)} \\
 + \frac{U' \sin \alpha \sin \xi_s \cos \xi_s \{ \cos \xi_s \sinh \eta_s \cos \alpha - \sin \xi_s \cosh \eta_s \sin \alpha \}}{\sinh \eta_s (\cosh \eta_s - \cos \xi_s) (\cosh^2 \eta_s - \cos^2 \xi_s)}.
 \end{aligned}$$

Leaving out the first term in the last expression, which is, of course, the velocity of the aerofoil along the line, the actual velocity of the vortex along the line is obtained.

The reduction of the two component velocities into the forms given in these expressions enable us to see immediately what order of magnitude they are. Since α is always in practice very small and consequently ξ_s is of the order π , then $\sin \alpha$ and $\sin \xi_s$ are both very small. U' may also be considered small and it is seen at once that in the first expression every term is of the order $\sin \alpha$, or higher and in the second expression every term is of the order $\sin^2 \alpha$ or higher. These expressions then give some justification for the assumption that for all practical purposes a single vortex may be assumed to remain at rest in the fluid. There is, however, in actual fact a first order movement sideways and a second order drift downstream.

Of course, having partly justified the assumption for a single vortex does not mean that it is quite justified when dealing with a trail of vortices coming off from a trailing edge. Each vortex of the trail will of course influence the others and will in turn be influenced by them. A perfectly straight trail with any distribution of vorticity along it will by itself produce no motion along its length—this is only produced by an image trail or such motions as calculated above. It will, however, produce a transverse motion, and if the trail is considered as being along the x -axis, and of strength m_s at the point x_s , such a transverse motion at the point x will be of the form

$$\int \frac{m_s dx_s}{2\pi(x_s - x)},$$

which has a proper value on the trail itself equal to

$$\frac{1}{2\pi} \int_{x_s}^{x-\epsilon} \frac{m_s dx_s}{x - x_s} - \frac{1}{2\pi} \int_{x+\epsilon}^x \frac{m_s dx_s}{x_s - x}.$$

Thus near one end, the trail will tend to move downwards, whilst at the other end it will tend to move upwards. The examination of this motion and the effect it has on the aerofoil will be considered in a future communication.

REFERENCES

- Durand, W. F. 1935 *Aerodynamic Theory*, 2, 301. Berlin.
 Karman, Th. von & Burgers, J. M. 1934 *Aerodynamic Theory*, 2, 102. Berlin.
 Karman, Th. von & Sears, W. R. 1938 *J. Aero. Sci.* 5, 379.
 Karman, Th. von & Tsien, L. C. 1945 *Quart. Appl. Math.* 3, 1.
 Kussner, H. G. & Schwarz, L. 1940 *Luftfahrtforschung*, 17, 355.
 Lamb, H. 1916 *Hydrodynamics*, 4th ed. p. 18.
 Morris, R. M. 1937 *Proc. Roy. Soc. A*, 161, 406.
 Morris, R. M. 1938 *Proc. Roy. Soc. A*, 164, 346.
 Morris, R. M. 1939 *Proc. Roy. Soc. A*, 172, 213.
 Schwabe, M. 1943 *Tech. Mem. Nat. Adv. Comm. Aeronaut.* no. 1039.
 Sears, W. R. 1940 *J. Franklin Inst.* 230, 95.
 Sears, W. R. 1941 *J. Aero. Sci.* 8, 104.
 Sohngen, H. 1940 *Luftfahrtforschung*, 17, 401.

Transverse magnetization in ferromagnetic crystals in relation to domain structure

BY A. VON ENGEL AND M. S. WILLS
Electrical Laboratory, Oxford University

(Communicated by E. C. Stoner, F.R.S.—Received 27 March 1946)

In the formal theory of ferromagnetic anisotropy the magnetic energy of a crystal is expressed in terms of the magnetic field strength and the angles between the easy directions of magnetization and the field and magnetization vectors respectively. Although the magnetization curves of single crystals are in good agreement with this theory as regards the components of magnetization parallel to the field, this is not true of the transverse component. In this paper the treatment is considered with special reference to the transverse component for the planes (100), (110) and (111). According to the existing theory the transverse component should increase as the field strength diminishes whereas the experimental results show that it becomes zero in very weak fields.

To account for this a simple treatment is here developed which covers the whole range of field strength. The directions of magnetization of domains are treated as distributed continuously in angle, rather than as restricted to a limited number of particular directions. By assuming that the proportion of the volume of domains magnetized in any direction is larger the lower the energy of magnetization in that direction, reasonable agreement with the experimental results is obtained.

Although the distribution assumed is of the Boltzmann type, the energy of distribution is not of thermal origin and its value is very much greater than kT . It is likely that the distribution is due to internal stresses or lattice imperfections.

The observations by Honda & Kaya have been newly corrected for demagnetization as regards magnitude and angle. The discussion is restricted by the errors in the experiment owing to demagnetization, non-uniformity of magnetization and preparation of the crystals.

1. INTRODUCTION

When a ferromagnetic single crystal is magnetized by a uniform magnetic field the direction of the magnetization is in general different from that of the field. With a given small field the crystal is most easily magnetized in certain directions, called directions of easy magnetization. The angle between field and magnetization vanishes for these directions in the crystal. The easy directions for body-centred cubic iron are the cube edges [100], for face-centred nickel the cube diagonals [111], and for the cobalt hexagonal prism the prism axis [0001].

The facts that such easy directions exist and that in ferromagnetics exceptionally large magnetizations are observed in very weak fields are explained by the domain theory of Weiss (Stoner 1934; Bitter 1937; Bates 1939; Becker & Döring 1939). Domains are small regions spontaneously magnetized to saturation containing on the average between 10^{10} and 10^{15} atoms; in general there is a large number of domains in a single crystal so that its structure can be regarded as homogeneous. The spontaneous magnetization has been shown to originate from electron spins which are aligned by exchange forces (Heisenberg 1928, 1931). The reason why spontaneous magnetization is confined to small regions is not quite clear. It has

been proposed, for instance, that the demagnetization field of a domain is one of the factors responsible for limiting its size (Becker & Döring 1939; Stoner 1944), but internal stresses also have a pronounced effect on the size of domains.

In an unmagnetized substance—according to this picture—the directions of magnetization of the domains are assembled along the axes of easy magnetization in such a way that the resultant magnetic moment of the substance is zero. In moving from one domain to a neighbouring one, a region is passed through where there is a more or less steep change of directions of spins; this region is generally termed a wall or boundary. The application of a small external field causes the spins at the boundary to turn so that one domain grows at the expense of its neighbour. This displacement of the boundary, known as a translational process, causes the domains to become magnetized, in general along the easy direction which is nearest to the field. As these changes of magnetization are only from one easy direction to another, without intermediate positions of equilibrium, it follows that the translational processes are associated with a turning of the spins through a large angle, 90° or 180° in the case of iron. These changes produce the Barkhausen effect when a large number of spins are affected, as a boundary moves spontaneously from one position of equilibrium to another.

If the field is along an easy direction, saturation magnetization is obtained in a very weak field; it is, however, still uncertain whether an ideal crystal would be saturated by an infinitesimal field. If the field is in another direction, it is observed that saturation to the same intensity can still be obtained if only a sufficiently strong field is applied. This makes it necessary to assume that under these conditions a gradual turning of the spins from an easy direction towards the field direction takes place; this is known as a rotational process.

As the translational processes are completed in a weak field, the magnetization of the crystal may be represented in all but the weakest fields by a vector of magnitude equal to I_0 , the saturation intensity. This vector thus lies for a small field in the easy direction nearest to the field; with increasing field the vector rotates towards the field direction reaching it at infinite field strength. We shall show in § 2 how the position of the vector for any given field is computed from energy considerations.

By the use of what will be called the 'single-vector treatment' indicated above it is possible to calculate the angular variation of the intensity of magnetization as a function of the field strength. The results may be conveniently expressed by resolving the magnetization vector I_0 into components I_p parallel to the field and I_n normal to the field.

Investigations on single crystals of ferromagnetic substances reveal agreement between the results of this treatment and experiments as far as the dependence of I_p on the direction and strength of the field is concerned, but only for fields of more than about 100 oersteds. For more moderate fields this theory does not meet with the same success. Again this theory explains the observations of the transverse component I_n in strong fields of more than about 1500 oersteds, but it fails completely in moderate and small fields. As I_p and I_n are only components of the same

vector, a good agreement of I_n alone can only be superficial, and the failure to explain the observations of I_n points to a fundamental weakness in the current theory.

The weakness of the single-vector treatment arises from the fact that the effects of domain structure are ignored. Such a structure must, however, be assumed to account at least qualitatively for the initial part of the magnetization curve in which translational processes take part. A treatment applicable to moderate fields must therefore employ a concept in which translational as well as rotational processes are involved.

It is the aim of this paper first to compute by the existing treatment the angular relation between the transverse magnetization and the field strength in greater detail than has hitherto been done for moderate fields; secondly, to develop an elementary treatment which gives better agreement with experiment in the range of moderate fields. Hysteresis effects will be excluded.

2. SINGLE-VECTOR TREATMENT

The potential energy of a crystal is expressed in the usual treatment by the sum of elastic and magnetic terms. In what follows only magnetic effects will be considered and the elastic energy terms will be neglected. Further, it will be assumed that the crystal is completely homogeneous and that its temperature is constant. Since the crystal is being treated here as uniformly magnetized as a single unit, the state of minimum free energy reduces effectively to the state of minimum potential energy. Hence if the crystal is subjected to a magnetic field, the resulting magnetization is found from the assumption that in equilibrium the potential energy becomes an absolute minimum.

Consider a cubic crystal and let the axes of reference (i, j, k) be along the cube edges. Suppose the crystal to be magnetized with uniform intensity I_0 , the direction cosines of I_0 being (a, b, c) . The energy per unit volume of the crystal as far as terms involving the fourth power of the direction cosines reduces, owing to the symmetry of a cubic crystal (Stoner 1934), to

$$E = K(b^2c^2 + c^2a^2 + a^2b^2) - I_0H \cos \psi, \quad (1)$$

where H is the true field strength, i.e. the field corrected for demagnetization,

$$a = \cos \angle(I_0, i) \text{ etc. and } \psi = \angle(I_0, H). \quad (1a)$$

Equation (1) is derived under the assumption that the magnetic energy consists of internal energy independent of H plus the energy of a dipole $-I_0H \cos \psi$. The expression used is in sufficiently close agreement with experiment for the present purpose, but it should be noted that K , the anisotropy constant, depends on temperature, and appears to increase slightly with H (Bitter 1937). The internal-energy term is the lowest order term which is relevant in the expression for the 'anisotropy' energy, whose form may be determined purely from symmetry considerations, but

not the values of the coefficients, which must be determined from experiment. For iron at room temperature

$$K = 4 \times 10^5 \text{ erg/cm.}^3 \quad \text{and} \quad I_0 \doteq 1720 \text{ e.m.u.}$$

The theory assumes that I_0 is a vector of constant magnitude which is supposed to exist already in an infinitesimal field, and that in a field of given strength and direction I_0 turns into the direction for which the value of E is a minimum. For some inclinations of H where there are two or more equal minima of E , the vector may be split into equal components along the several directions. Physically this means that equal volumes of domains are magnetized along these directions. In sufficiently strong fields one minimum coinciding with the field predominates.

Using these ideas Akulov (1929, 1931) has, for example, computed the I_p - H curves for single crystals of iron for the [100], [110] and [111] directions. The agreement is very good except for fields below 100 oersteds and between 350 and 500 oersteds for the [111] direction.

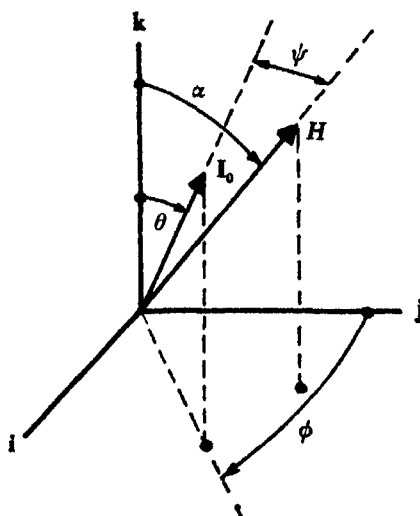


FIGURE 1. Definition of notation.

The normal component I_n can be computed as follows. From (1), changing to polar co-ordinates as shown in figure 1,

$$E = \frac{K}{8} [\frac{1}{2}(3 - 4 \cos 2\theta + \cos 4\theta)(1 - \cos 4\phi) + (1 - \cos 4\theta)] - I_0 H \cos \psi, \quad (2)$$

$$E = E_1 - I_0 H \cos \psi. \quad (3)$$

It will be shown that the normal component with the field vector in the (100) plane behaves anomalously, a fact which has hitherto escaped attention. For the field in the (110) and (111) planes the normal component is calculated systematically for the first time and similar anomalies are again found. These peculiarities occur,

of course, in medium fields only, and it is not certain whether their existence can be confirmed by future experiments because they lie in the range where the single-vector treatment begins to fail.

(a) *The (100) plane*

Let \mathbf{H} be in a (100) plane at an angle α with \mathbf{k} . From the symmetry it follows that the absolute minimum of E must also lie in the (100) plane hence, putting $\phi = 0$ and $\psi = \alpha - \theta$ (see figure 1),

$$E/K/8 = (1 - \cos 4\theta) - 8h \cos(\alpha - \theta), \quad (4)$$

where the relative field strength

$$h = I_0 H/K. \quad (5)$$

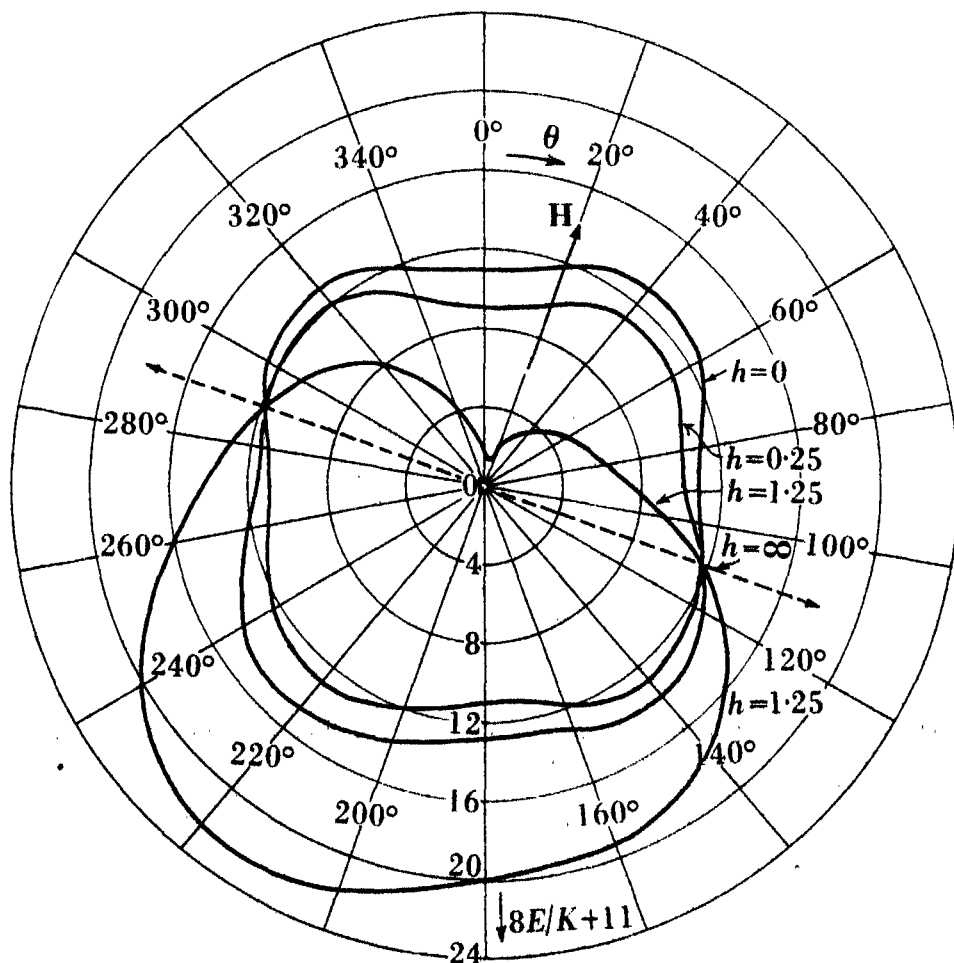


FIGURE 2. Relative values of the energy of magnetization of a single iron crystal in a (100) plane as a function of the angle θ for different values of the relative field strength h ; \mathbf{H} inclined at $\alpha = 20^\circ$.

With the above given values of I_0 and K , $h = 1$ is found to correspond to $H \approx 228$ oersteds. The stationary values of E are given by $dE/d\theta = 0$, viz.

$$\sin 4\theta = 2h \sin(\alpha - \theta). \quad (6)$$

This equation solved graphically gives all the maxima and minima, and it is necessary to select the absolute minimum by drawing curves of E against θ . For convenience to avoid negative values the quantity $11 + E/K/8$ has been used. In figure 2 curves for $\alpha = 20^\circ$ and $h = 0, 0.25, 1.25$ and ∞ are given which show how the energy distribution of the crystalline forces is gradually distorted by an applied field. As the field increases the original energy distribution, symmetrical with respect to the three axes, is gradually changed into a distribution symmetrical about the field direction. The energy perpendicular to the field is of course unchanged, and hence the curves intersect at $\theta = 110^\circ$ and 290° . In sufficiently large fields the crystalline forces are no longer important, and the polar diagram of the energy distribution tends towards the circle $11 + E/K/8 = -8h \cos(\alpha - \theta)$ which becomes of infinite radius at $h = \infty$.

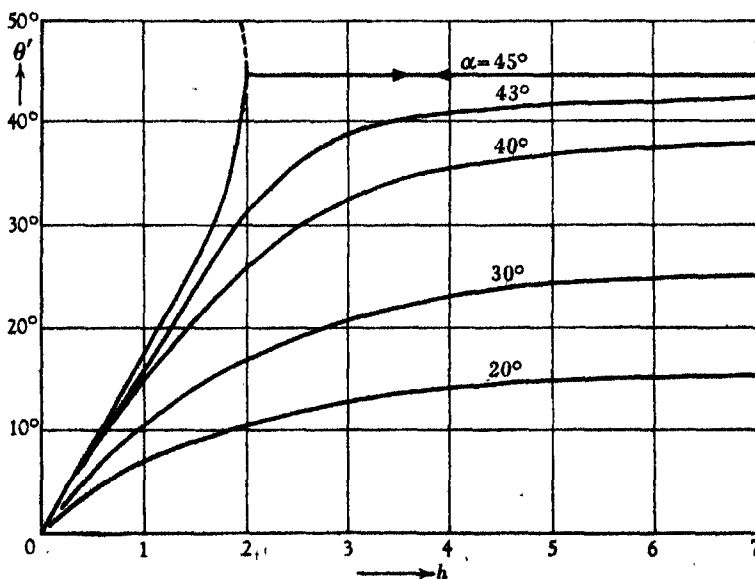


FIGURE 3. Angular position of the relative energy minima in the (100) plane as derived from (6) as a function of the relative field strength for different values of α .

The value of θ for which E has its lowest minimum value will be denoted by θ' . Figure 3 shows curves of θ' against h for different values of α as parameter. In the range $\alpha = 0$ to 45° θ' increases from zero at $h = 0$ asymptotically to $\theta' = \alpha$ at $h = \infty$, except for $\alpha = 45^\circ$. This is in accordance with the fact that in an infinitesimal field I_0 is along k and tends to turn gradually into the direction of H as it increases. In the range $\alpha = 45-90^\circ$ the curves are symmetrical with respect to the line $\alpha = 45^\circ$.

With these values of θ' the relative value of the normal component given by

$$i_n = \frac{I_n}{I_0} = \sin(\alpha - \theta') \quad (7)$$

may be calculated. In figure 4, i_n is plotted as a function of α for h between 0 and 5. For small values of h the curve approximates to part of a sine curve with discontinuities and the maximum value of i_n at $\alpha = 45^\circ, 135^\circ$, etc. When h is increased the maxima become smaller and shift towards $\alpha = 22.5^\circ, 67.5^\circ$, etc., and the discontinuity eventually disappears. There are two critical values of h , namely, the value at which the maximum of i_n begins to move away from $\alpha = 45^\circ$, and the value at which discontinuities disappear from the curve. These field strengths have been calculated to be $h = 1.307$ and $h = 2.0$ respectively. According to (5), for iron, $h = 1$ is equivalent to a field strength of about 230 oersteds.

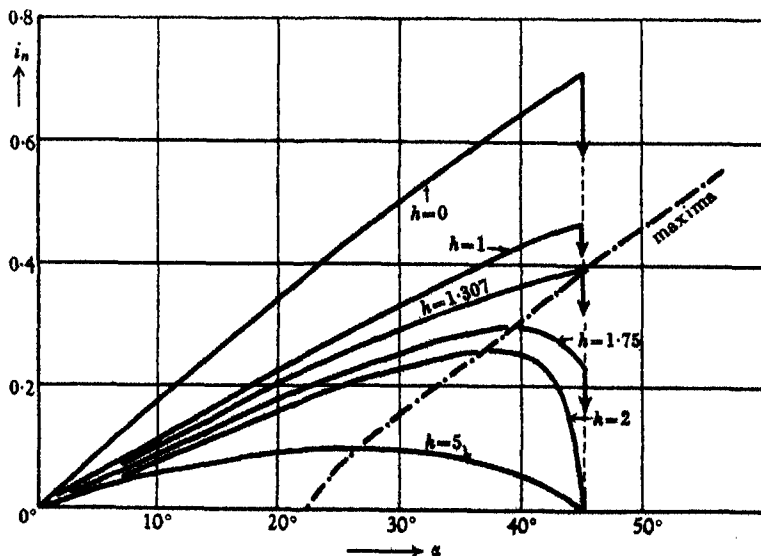


FIGURE 4. Transverse magnetization (in units of I_0) as a function of the field direction in the (100) plane for different values of the relative field strength as derived from single vector treatment. Honda & Kaya (1926) obtained only continuous curves of 90° periodicity.

Though, as will be seen later, this theory fails altogether in weak fields it is still uncertain, owing to factors discussed in § 3, at what h the deviations begin to be appreciable and how far the above results are in agreement with the facts.

Another way of representing the results of this treatment is to eliminate α between (6) and (7),

$$i_n = \frac{1}{2h} \sin 4\theta' = \frac{K}{2I_0 H} \sin 4\theta'. \quad (8)$$

Honda & Okubo (1916) have first examined (8), and later Gans (1933) and others have used some experimental values of Honda & Kaya (1926) to test this result for large values of H and found it confirmed.

(b) The (110) plane

Now suppose that \mathbf{H} is confined to a (110) plane as shown in figure 5. When $\alpha = 0$ the lowest minimum of E is clearly in the \mathbf{k} direction, and as α is increased up to 54.7° the minimum lies in the (110) plane between \mathbf{k} and \mathbf{H} directions. When \mathbf{H} is along the trigonal axis ($\alpha = 54.7^\circ$) it follows from symmetry that there are three minima of equal depth. When $\alpha > 54.7^\circ$ the two minima near to the \mathbf{i} and \mathbf{j} axes are deeper and the vector of magnetization is split into two components $\frac{1}{2}\mathbf{I}_0$ (figure 5, dash-dotted) along the directions of these two minima; this splitting is peculiar to \mathbf{H} being exactly in the (110) plane.

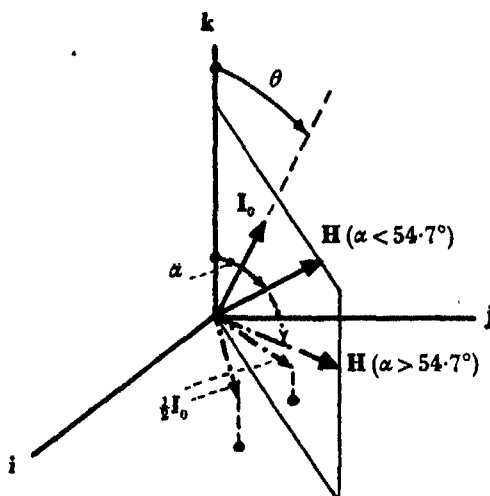


FIGURE 5. Vector of magnetization when \mathbf{H} lies in the (110) plane.

The i_n curves have been computed as follows. The direction cosines of \mathbf{H} are $\{(\sin \alpha)/\sqrt{2}, (\sin \alpha)/\sqrt{2}, \cos \alpha\}$ and those of \mathbf{I}_0 are $\{\sin \theta \sin \phi, \sin \theta \cos \phi, \cos \theta\}$ from which $\cos \psi$ is deduced. Then, from (1) and (2),

$$E/K/8 = \frac{1}{8}(3 - 4 \cos 2\theta + \cos 4\theta)(1 - \cos 4\phi) + (1 - \cos 4\theta) - 8h \left[\frac{\sin \alpha \sin \theta}{\sqrt{2}} (\sin \phi + \cos \phi) + \cos \alpha \cos \theta \right]. \quad (9)$$

The stationary values from which the lowest minima are to be selected are given by

$$\frac{\partial E}{\partial \theta} = \frac{\partial E}{\partial \phi} = 0. \quad (10)$$

Differentiating (9), then

$$\frac{8 \partial E}{K \partial \theta} = \{ \frac{1}{2}(2 \sin 2\theta - \sin 4\theta)(1 - \cos 4\phi) + 4 \sin 4\theta \} - 8h \left\{ \frac{\sin \alpha \cos \theta}{\sqrt{2}} (\sin \phi + \cos \phi) - \cos \alpha \sin \theta \right\}, \quad (11)$$

$$\frac{8 \partial E}{K \partial \phi} = \{ \frac{1}{2}(3 - 4 \cos 2\theta + \cos 4\theta) \sin 4\phi \} - 8h \left\{ \frac{\sin \alpha \sin \theta}{\sqrt{2}} (\cos \phi - \sin \phi) \right\}. \quad (12)$$

The values of θ and ϕ corresponding to the lowest minima are denoted by θ' and ϕ' . The correct values of θ' and ϕ' have been found in this case by successive approximation so as to satisfy both (11) and (12) simultaneously, and i_n has been calculated from

$$i_n = \cos \theta' \sin \alpha - \sin \theta' \cos (45 - \phi') \cos \alpha. \quad (13)$$

When $0 < \alpha < 54.7^\circ$ and the minimum lies in the (110) plane $\phi' = 45^\circ$ and (13) reduces to (7).

Again if $\phi' = 45^\circ$ then from (11) and (7)

$$i_n = (2 \sin 2\theta' + 3 \sin 4\theta')/8h, \quad (14)$$

$$\text{which approximates to } i_n \doteq (2 \sin 2\alpha + 3 \sin 4\alpha)/8h \quad (15)$$

if only h is large enough.

From (13) the relation between i_n and α is plotted for different values of h in figure 6. For small values of h , i_n rises with increasing α to a maximum at 54.7° and then changes over discontinuously to negative values. When $h > 1.15$ the maxima occur at smaller values of α and a discontinuity still exists at 54.7° . At $h = 1.86$ the discontinuity disappears and, for higher values of h , i_n is given approximately by (15).

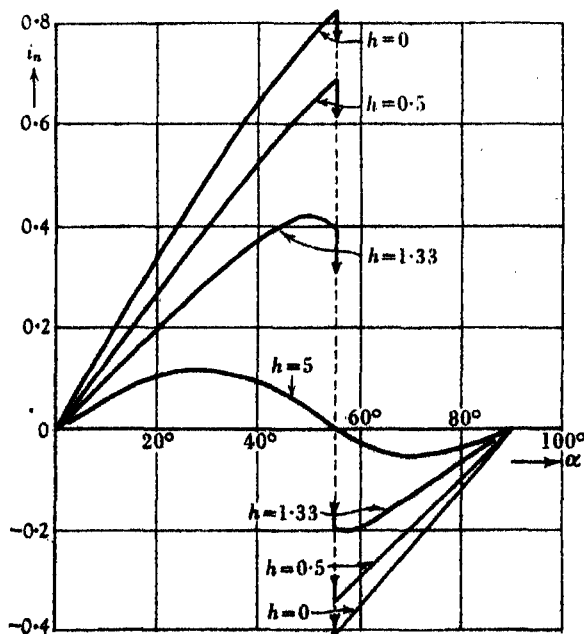


FIGURE 6. Transverse magnetization (in units of I_0) as a function of the field direction in the (110) plane for different values of the relative field strength as derived from single-vector treatment. Honda & Kaya (1926) obtained only continuous curves of 180° periodicity.

(c) The (111) plane

Finally, suppose \mathbf{H} to be confined to a (111) plane at an angle γ with the \mathbf{k} axis as shown in figure 7. The direction cosines of \mathbf{H} are

$$\left\{ \frac{1}{\sqrt{2}}[-\sin \gamma - (\cos \gamma)/\sqrt{3}], \quad \frac{1}{\sqrt{2}}[\sin \gamma - (\cos \gamma)/\sqrt{3}], \quad \sqrt{\frac{2}{3}} \cos \gamma \right\},$$

and of \mathbf{I}_0 as before, $\{\sin \theta \sin \phi, \sin \theta \cos \phi, \cos \theta\}$

and from (3) the energy becomes

$$\frac{8E}{K} = E_1 - 8h \left\{ \frac{\sin \theta \sin \gamma}{\sqrt{2}} (\cos \phi - \sin \phi) - \frac{\sin \theta \sin \gamma}{\sqrt{6}} (\cos \phi + \sin \phi) + \sqrt{\frac{2}{3}} \cos \theta \cos \gamma \right\}. \quad (16)$$

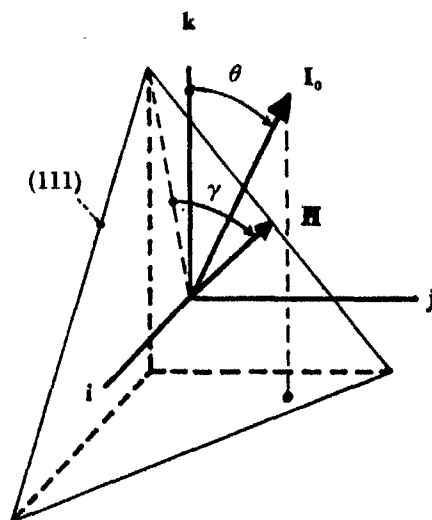


FIGURE 7. Vector of magnetization when \mathbf{H} lies in the (111) plane.

As before the stationary values satisfy (10) and the values θ' , ϕ' for the lowest minima have been selected and i_n calculated from

$$i_n = \sin \angle (\mathbf{H}, \mathbf{I}_0), \quad (17)$$

where in this case

$$\cos \angle (\mathbf{H}, \mathbf{I}_0) = \frac{\sin \theta \sin \gamma}{\sqrt{2}} (\cos \phi - \sin \phi) - \frac{\sin \theta \cos \gamma}{\sqrt{6}} (\cos \phi + \sin \phi) + \sqrt{\frac{2}{3}} \cos \theta \cos \gamma. \quad (18)$$

Figure 8 shows i_n plotted against γ for different values of h . The angular position of \mathbf{H} has been expressed by the angle γ because this, and not the angle α , is readily measured experimentally, and because besides α another angle would have to be introduced to define the direction of \mathbf{H} relative to the three axes.

In figure 8 a certain number of points at $\gamma = 0^\circ, 30^\circ$, etc., are identical with corresponding ones on the (100) and (110) curves; e.g. the direction $\gamma = 0$ is identical with

$\alpha = 35.3^\circ$ in a (110) plane. As before discontinuities occur, at 30° and 60° , which disappear in large fields.

It should be noted that i_n as given by (17) is not in general in the (111) plane, and that in some experiments the measured quantity is the component of this vector in the plane. This explains why the 120° periodicity of figure 8 differs from the 60° periodicity measured by Honda & Kaya (1926).

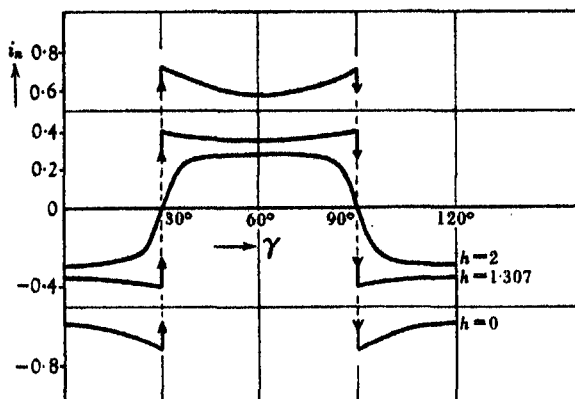


FIGURE 8. Transverse magnetization (in units of I_0) as a function of the field direction in the (111) plane as defined by figure 7 for different values of the relative field strength as derived from single-vector treatment.

3. STATISTICAL TREATMENT OF DOMAIN EFFECTS

In figure 9 observations of the relative transverse magnetization in single iron crystals by Honda & Kaya (1926) and Webster (1925) are plotted as a function of the relative field strength for an angle $\alpha = 40^\circ$. The same figure contains the curve derived from (8). The observations show that in strong fields i_n is small and increases as h decreases up to a maximum, after which it decreases and appears to tend towards zero as h approaches zero. The theoretical curve agrees closely with observations in strong fields, i.e. $h > 10$ or $H > 2000$ oersteds. As the field is reduced, however, it diverges increasingly from the experimental curve and reaches a maximum value 0.64 at $h = 0$. Although doubt has been expressed as to the interpretation of Honda's measurements, a point which will be enlarged upon in § 4, corrections do not seem to change the trend of this curve, and the marked divergence between theory and experiment remains. Similar divergencies are found in general for other orientations of the field.

There have been few attempts to explain this difficulty, the most noteworthy being that of Bozorth (1932). Using a method similar to that in Heisenberg's calculation of magnetostriction, Bozorth assumes that at zero field strength $\frac{1}{6}$ of all the domains, supposed of constant volume and variable in number, assemble along the six easy directions and in a finite field a redistribution along the six directions takes place so that the most probable distribution producing a given $i_p = I_p/I_0$ is

achieved. This is in contradistinction to the single-vector treatment of § 2, according to which in a field all the domains are in general aligned along the easy direction nearest to the field direction. Bozorth points out that his method allows calculation of i_n , not as a function of the field strength, but as a function of i_p only. It is possible to derive from Bozorth's paper the relation between i_n and i_p which is shown in figure 10 again for $\alpha = 40^\circ$ together with Honda's observations. There is clearly a considerable difference between Bozorth's treatment and the uncorrected measurements. Figure 10 also contains the observations as corrected by Bozorth by a method which will be discussed in § 4. This improves the agreement although characteristic deviations remain.

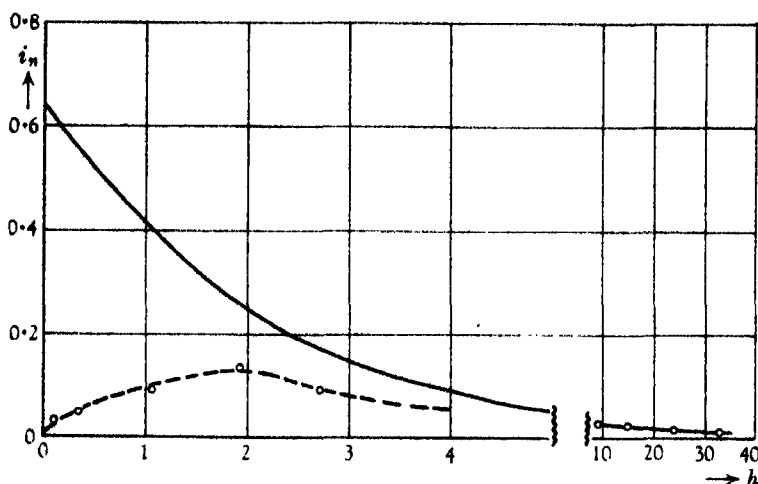


FIGURE 9. Transverse magnetization (in units of I_0) as a function of the relative field strength. Comparison between uncorrected observations and single-vector treatment for a field vector in the (100) plane, inclined at $\alpha = 40^\circ$ ($h = 1 \dots$, $H = 228$ oersteds). $-\circ-\circ-$ Observations by Webster (1925) on blast furnace crystals, and $-----$ by Honda & Kaya (1926) on crystals produced by the strain annealed process. $-----$ Single-vector treatment.

One of the weaknesses of this treatment is that it only refers to translational processes and hence, as may be seen from Bozorth's distribution curves, is intrinsically limited in scope to intensities of magnetization up to about 1100–1300 e.m.u., depending on the direction, and actually begins to fail at lower intensities.

Thus a gap remains between the completion of the translational processes and the region of the single-vector treatment which Bozorth assumes to become effective for larger values of i_p . It is noteworthy that for $i_p < 0.6$ the observed values of i_n and hence the corrected values are zero, while the theoretical curve approaches zero slowly as i_p becomes zero. One of the requirements of a new treatment is that it should give the dependence $I = f(H)$ without a break over the whole range of H .

The idea of an angular distribution of the directions of magnetization of domains in place of a concentration in the direction corresponding to minimum potential energy seems to be one which will explain the decrease of i_n with decreasing field

strength in weak fields. On the other hand, a new treatment should include the single-vector treatment which is amply supported by experiment.

It is therefore proposed, that instead of a concentration of domains in the direction of the minimum crystalline energy or an axial distribution of domains of constant volume and variable in number, to choose a distribution of the form

$$dv_\theta = C \exp(-E/E_0) d\omega, \quad (19)$$

where dv_θ is the volume of domains in a solid angle $d\omega$ orientated in a direction θ , E is the energy of unit volume of the crystal when magnetized in θ, ϕ -direction (see figure 1), and E_0 is a constant reference energy, the significance of which will be discussed below.

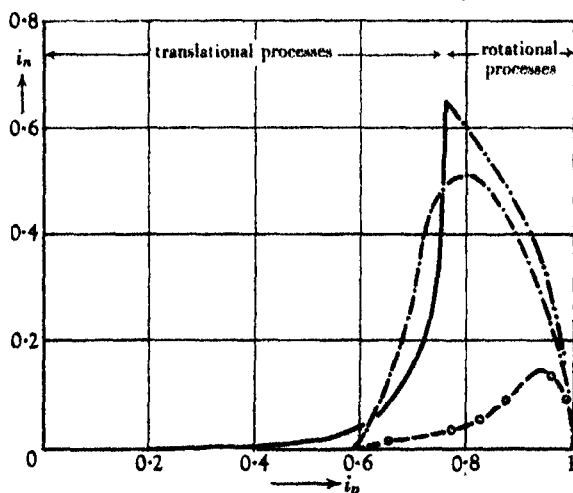


FIGURE 10. Transverse magnetization as a function of the longitudinal magnetization. Comparison between single-vector treatment \cdots , Bozorth's treatment \cdots and the corrected \cdots and uncorrected $-\circ-$ observations by Honda & Kaya (1926).

In figure 11 is shown a section through this distribution in a (100) plane for an unmagnetized crystal between $\theta = \pm 45^\circ$ for $E_0 = K/8$, from which it may be seen that for this value of E_0 there is still a certain volume of domains at angles up to 45° with easy directions. For example, the volume of domains per unit solid angle at 45° is 13.5 % of that at 0° —the easy direction. In the new treatment in an unmagnetized crystal the axes are merely directions of greatest probability for domains and the application of a field shifts the direction of greatest probability towards the field.

Figures 12a and b show the difference between the old assemblage and the new domain distribution in an unmagnetized cubic crystal, figures 12c and d the same in a weak field along the cube diagonal; figures 12e and f represent the domain distribution for different values of E_0 . The value of E_0 determines the strength of the crystalline forces in relation to that of the forces causing the distribution. If E_0 is

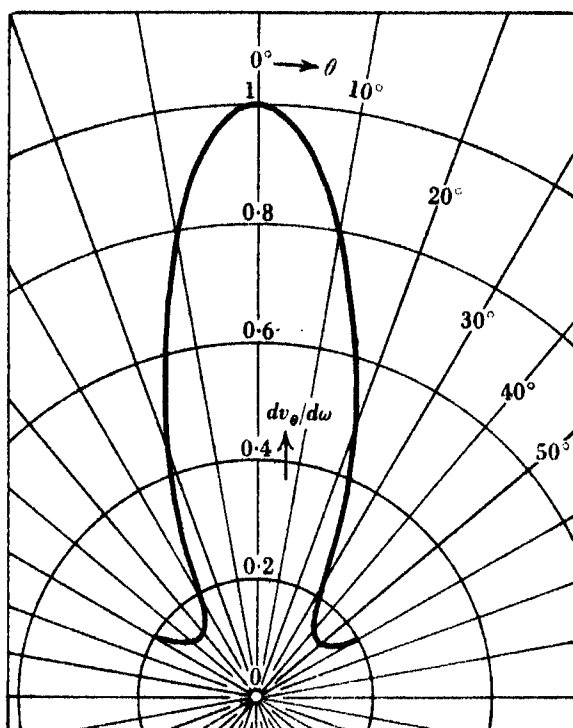


FIGURE 11. Angular distribution of domain volume in a (100) plane at $h=0$ for an energy of distribution $E_0 = K/8$.

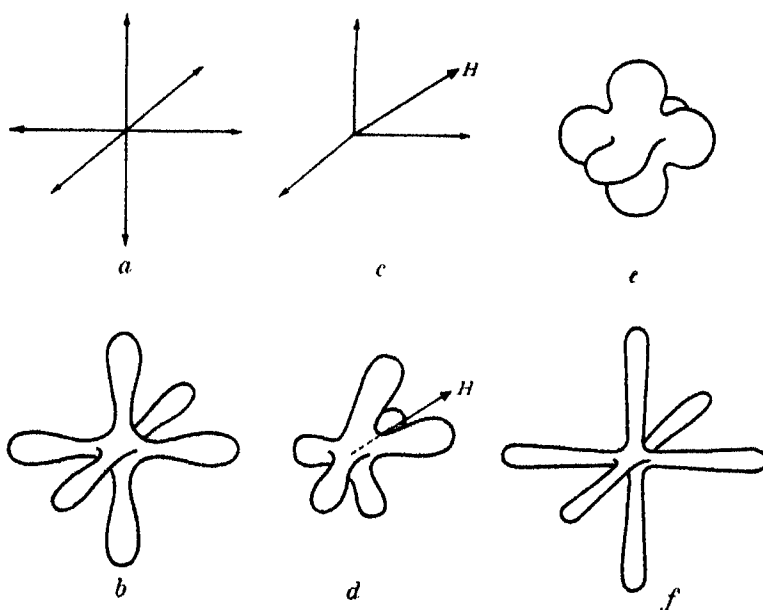


FIGURE 12. Classical axial assemblage and solids representing the new distribution of directions of domains (all solids have equal volume). a and b , $H=0$; c and d , weak H ; e and f , $H=0$, large and small E_0 respectively.

rather large the distribution is as indicated in figure 12*e*, and if small as in figure 12*f*. It is to be understood that the total volume of domains represented by any of the diagrams of figure 12 is the same. As $K \rightarrow \infty$, or $E_0 \rightarrow 0$, the distribution becomes narrower and tends in the limit to the axial assemblage of the old theory. If, on the other hand, $K \rightarrow 0$ or $E_0 \rightarrow \infty$, the substance loses its crystalline properties and the distribution is isotropic.

It will be convenient to consider the consequences of (19), treated as a purely formal hypothesis, before considering the problem of its physical interpretation.

It is possible by using (19) to calculate i_p and i_n for any direction of the field strength in the crystal, but the integrals become rather cumbersome. From (19) and the definition of I_p given in § 1 the value of i_p is found to be

$$i_p = \frac{I_p}{I_0} = \frac{\int \cos \psi dv_\theta}{\int dv_\theta} = \frac{\int \cos \psi \exp(-E/E_0) d\omega}{\int \exp(-E/E_0) d\omega}, \quad (20)$$

where E is the energy given by (3) in a general direction which makes an angle ψ with the field and $d\omega$ is an elementary solid angle around that direction. The value of i_n is given by

$$i_n = \frac{I_n}{I_0} = \frac{\int f(\theta, \phi) \exp(-E/E_0) d\omega}{\int \exp(-E/E_0) d\omega}. \quad (21)$$

The form of $f(\theta, \phi)$ depends on the direction of \mathbf{H} in the crystal. If, for example, \mathbf{H} is confined to a (100) plane,

$$f(\theta, \phi) = \cos \theta \sin \alpha - \sin \theta \cos \phi \cos \alpha. \quad (22)$$

To avoid evaluating these integrals the following approximate method was devised: instead of distributing the domains with a probability given by (19) over the whole solid angle, the simplifying assumption has been made that the domains are assembled in the directions of different energy minima according to the relative values of their potential energies. Thus in weak fields there are six nearly equal minima of energy, representing directions in which the total volume of domains is likely to distribute in nearly equal parts. When the field is increased (see figure 2) it has been shown that the minima nearer to the field become lower and the others higher; at the same time the angular position of the minima is so changed that the minima next to the field shift towards it. According to this view more domain volume is assembled in the deeper minima at the cost of the higher ones which finally become ineffective. Hence in stronger fields only a smaller number of minima need be considered in an approximate calculation. In sufficiently strong fields only one minimum remains, and thus the treatment goes over into the single-vector treatment.

The changes described above affect i_n in the following manner: in weak fields the distribution remains almost symmetrical with respect to H , and hence the transverse

FIGURE 13. Transverse magnetization as a function of the longitudinal magnetization (both in units of I_0) with H in the (100) plane for $\alpha = 20^\circ$ as derived from statistical treatment for different values of the energy of distribution ($E_0 = 0$ identical with single-vector treatment, $E_0 = \infty \dots$ isotropic medium) —○—. Observations by Honda & Kaya.

included. It would clearly be possible, by using a revised value of E_0 , to obtain closer agreement with the experiment, but owing to the rather large uncertainties in connexion with the experimental results, as well as the difficulties in connexion with the physical significance of E_0 , to be discussed in the next section, it was felt that more extensive detailed calculations would not at the present stage be helpful.

Similar results have been obtained for other values of α . It may be remarked that the position of the theoretical maximum of i_n is, in all cases investigated, close to the observed maximum and that i_n is negligibly small for $I_p \leq 900$ e.m.u.

In figure 14 the same comparison is made for the (110) plane except that the calculation refers to $\alpha = 45^\circ$, whereas the only observations available are for $\alpha = 40^\circ$.

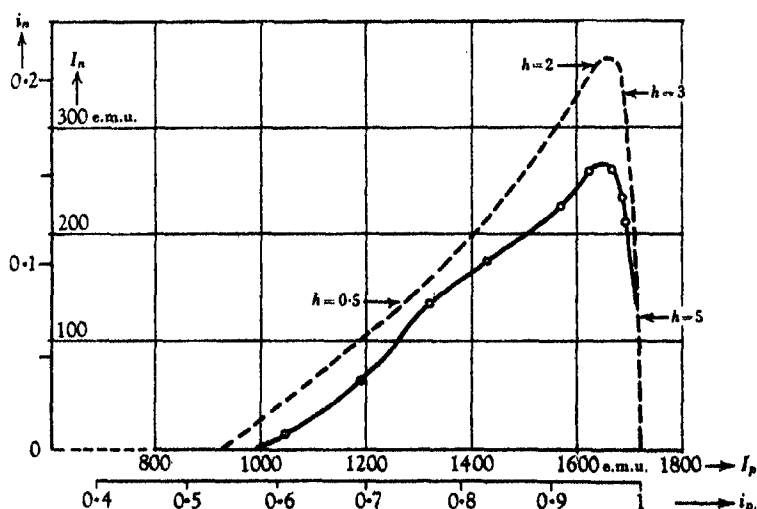


FIGURE 14. Transverse magnetization (in units of I_0) as a function of the longitudinal magnetization as derived from the statistical treatment. H in the (110) plane, $\alpha = 45^\circ$ (dashed curve; uncorrected observations by Honda & Kaya for $\alpha = 40^\circ$ (full curve).

Treating E_0 as a parameter to be determined from the experimental results, it is not possible to determine the value with precision. It may, however, be said that the experimental results which have been considered indicate a value of E_0 lying between the limits $K/8$ and $\frac{1}{4}K/8$, demagnetization being allowed for (see below). This is shown by using the corrected values of i_n (figure 16) in figure 13, where i_n is plotted against i_p for different values of E_0 between zero and infinity.

The true field in all known cases is derived from the applied field by subtracting a quantity for demagnetization. In moderate fields this quantity is of the same order as the applied field so that a large error may be involved. As the vector of demagnetization is parallel to I_0 which is in general not parallel to H , the direction of the true field is different from that of the applied field. Figure 15 shows how demagnetization can be allowed for. H_a is the applied field and I'_p and I'_n the observed parallel and transverse components of magnetization. These give rise to

the demagnetization components DI'_p , DI'_n , from which the magnitude and direction of the resulting field is found. The components of I_0 along and perpendicular to H are given by

$$I_p = I'_p \cos \delta - I'_n \sin \delta, \quad (23)$$

$$I_n = I'_p \sin \delta + I'_n \cos \delta, \quad (24)$$

where δ is the angle between H and H_a . The magnitude and direction of H are given by

$$H = (H_a - DI'_p) \sec \delta, \quad (25)$$

$$\tan \delta = DI'_n / (H_a - DI'_p). \quad (26)$$

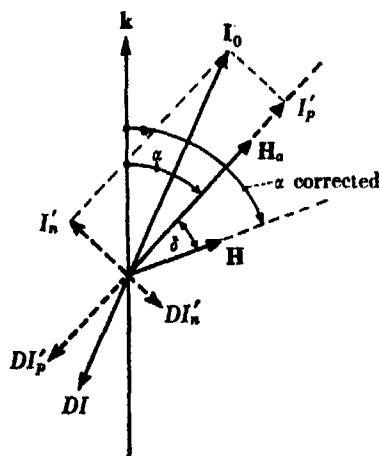


FIGURE 15. Vector diagram related to the correction for demagnetization.

By use of these equations the observations by Honda & Kaya have been corrected and the results are plotted in figure 16. D in (25) was of order 0.2. Clearly, as in the observations the applied field was kept constant, a complete set of curves for definite values of H can only be obtained approximately by interpolation. The general character of the curves is not affected by the correction, but the maximum values of i_n are perhaps three times as large. A great many points are found lying near $\alpha = 45^\circ$ where the expected intensity would be negligibly small, and this seems to indicate some error in angular measurements. Figure 16 shows that an angular error up to 5° in α might be involved which might have been due to inaccurate location of the crystal axis or to error in the coefficient of demagnetization; the latter can vary if the magnetization is non-uniform. Broadly speaking, if allowance for this error is made, the final values of i_n are all smaller in magnitude, i.e. the original correction has been exaggerated.

Honda & Kaya (1926) corrected the magnitude of H but did not allow for angular changes. Webster (1925) mentioned all the required corrections but did not apply them. Only Bozorth (1932) corrected for both factors but did not mention the above discrepancy.

Bozorth also suggests that errors may have been introduced into Honda & Kaya's experiments by turning of the specimen resulting in rotational hysteresis. This effect has been observed by von Harlem (1932) in experiments on polycrystalline specimens. Rotation of the specimen causes angular displacement of the magnetization vector in the direction of rotation which is independent of the speed of rotation and depends on H in such a way that it becomes zero for small and large H . This effect, however, seems to be closely related to normal hysteresis. Since in single crystals normal hysteresis is in general negligibly small (Honda & Kaya 1926), it is not to be expected that rotational hysteresis plays an appreciable part.

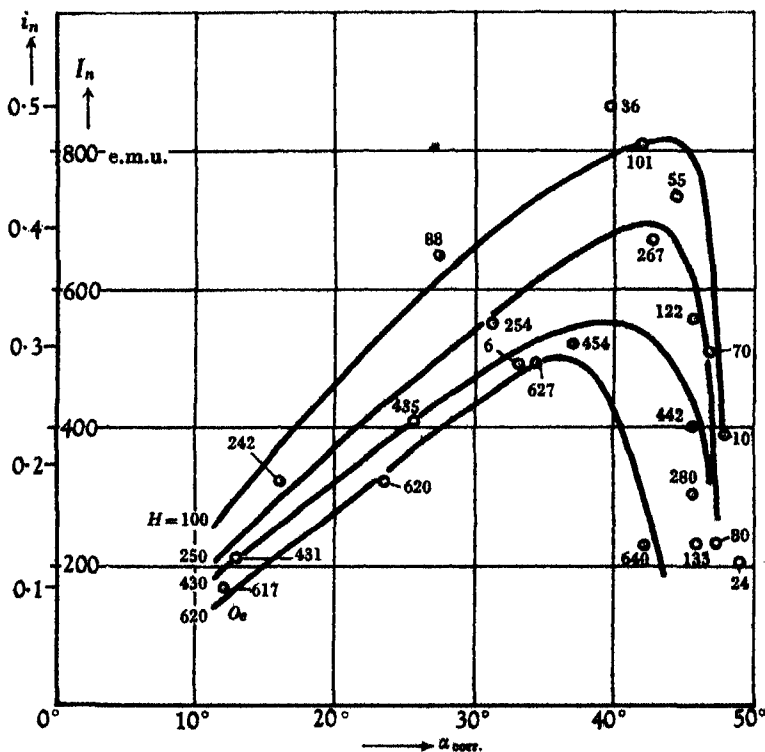


FIGURE 16. Corrected values of the transverse component as a function of the corrected angle α (see figure 15) for different values of H . The figures denote the corrected field strength.

5. DISCUSSION

It is necessary to discuss the choice of E_0 , the energy distribution parameter, an increasing value of which broadens the distribution of directions of domains.

The assumption of a distribution though it is formally identical with the Boltzmann distribution corresponding to (19) does not imply that the energy of distribution is of thermal origin, i.e. of the order kT/\bar{v} , where \bar{v} is the average volume of a domain; in fact E_0 has to be taken much larger than the thermal energy by a

factor 10^8 , perhaps, if the distribution is to explain the observations. Putting $E_0 = K/8 = 5 \times 10^4$ ergs/c.c. with $T = 300^\circ \text{K}$, $k = 1.37 \times 10^{-16}$ erg/ $^\circ \text{K}$, the corresponding value of \bar{v} is about 10^{-18} c.c., a volume corresponding to about 8×10^4 atoms for iron.

This is much smaller than the sizes usually attributed, with good experimental backing, to domains, namely, 10^{10} – 10^{15} atoms or volumes roughly from 10^{-13} – 10^{-8} c.c. from which it can be concluded that the temperature, i.e. the energy of agitation in the classical sense, has no direct bearing on the problem apart from the effects which occur near the Curie point.

Though there is at present no definite answer to the physical meaning of E_0 , one possible explanation is that it is related to the existence of lattice imperfections and associated stresses which have the required disorganizing influence on the regularity with which the directions of magnetization are distributed throughout the domains. These imperfections are known to depend to a large extent on the method of growing the single crystals, and it is uncertain whether they can be avoided at all. The perfection of a single crystal is also hampered by impurities which cause stresses during its growth, and dissolved gases have a similar effect.

Most of the relevant magnetic measurements on single iron crystals have been carried out at a time when secondary structures have not been known. However, monocrystalline specimens consist usually of a mosaic of small crystalline blocks (W. H. & W. L. Bragg 1933) which are inclined at small angles to each other. These angles are probably of the order of 1 min. as found by X-ray measurements. The mosaic structure plays an important part in the theory of plastic deformation of crystals (Taylor 1934). In this theory a length is defined which represents 'the distance through which a dislocation can travel freely along a slip-plane before being held up by some fault in the perfection of the regular crystal structure'. It represents therefore a linear dimension of a cell of a mosaic. The length of a side of such a mosaic has been derived from plastic stress-strain curves. It amounts for iron crystals to 1.7×10^{-4} cm. On the other hand, Belaiew (1925) by etching iron crystals observed a spacing of surface marks to be 0.25×10^{-4} cm. With these values we have mosaic volumes of 5×10^{-12} and 1.6×10^{-14} c.c. This compares favourably with the observed volume of a domain of 10^{-8} – 10^{-13} c.c.

From the observed mean deviation of 1 min. between the axes of the mosaic block and the axis of a single crystal an estimate of the energy has been made (cf. Schmid & Boas 1935) which is necessary to produce plastic deformations and hence to release the internal stress. This leads to a value of 10^4 ergs/c.c., which is of the same order as the value of E_0 required by experiments. This, of course, may only be a coincidence.

Thus it appears as if the blocks of the mosaic could be identified with the magnetic domains. One objection is that according to our assumption (see figure 11) an appreciable number of domains is expected to be assembled in a direction of 45° inclined against the axis, whereas only 1 min. deviations are observed with X-rays. However, it has to be borne in mind that the X-ray measurements refer to reflexions

from the planes of the blocks, while the direction of magnetization of a block ultimately depends on the distribution of directions of the electron spins throughout a domain.

After this investigation had been concluded, a paper by M. Takagi (1939) was found, dealing with statistical domain theory of ferromagnetic crystals. He obtains integrals of the form of (20) and by means of different approximations deduces formulae for the parallel component. The transverse component is derived from the relation $di_p/d\alpha = d(i_n h)/dh$ which seems to be a consequence of the exponential form assumed for the distribution. By using the observed values of i_p by Honda & Kaya and with $E_0 = K/20$ and after correcting for demagnetization, Takagi obtains a relatively good agreement for i_n in a (100) plane; this, he thought, could perhaps be improved by reducing E_0 to $K/70$.

The authors wish to express sincere thanks to Professor E. C. Stoner, F.R.S., for his encouragement, for reading the manuscript and making many valuable suggestions, to Professor Lord Cherwell, F.R.S., for stimulating discussions, to R. T. Lattey, T. C. Keeley and Professor F. Soddy, F.R.S., for providing facilities for carrying out the research.

REFERENCES

- Akulov, N. 1929 *Z. Phys.* **57**, 249.
 Akulov, N. 1931 *Z. Phys.* **69**, 78.
 Bates, L. F. 1939 *Modern magnetism*. Camb. Univ. Press.
 Becker, R. & Döring, W. 1939 *Ferromagnetismus*. Berlin: J. Springer.
 Belaiew, N. T. 1925 *Proc. Roy. Soc. A*, **108**, 295.
 Bitter, F. 1937 *Introduction to ferromagnetism*. London: McGraw Hill.
 Bozorth, R. M. 1932 *Phys. Rev.* **42**, 883.
 Bragg, W. H. & Bragg, W. L. 1933 *The crystalline state*, p. 214. London: G. Bell & Sons.
 Gans, R. 1933 Article in *Magnetismus* (Leipzig lectures), edited by P. Debye, p. 91ff. Leipzig: S. Hirzel.
 Harlem, J. von 1932 *Ann. Phys., Lpz.*, **14**, 667.
 Heisenberg, W. 1928 *Z. Phys.* **49**, 619.
 Heisenberg, W. 1931 *Z. Phys.* **69**, 287.
 Honda, K. & Kaya, S. 1926 *Sci. Rep. Tōhoku Univ.* **15**, 721.
 Honda, K. & Okubo, J. 1916 *Sci. Rep. Tōhoku Univ.* **5**, 153.
 Schmid, B. & Boas, W. 1935 *Kristallplastizität*, p. 290. Berlin: J. Springer.
 Stoner, E. C. 1934 *Magnetism and matter*. London: Methuen & Co.
 Stoner, E. C. 1944 Kelvin Lecture. *J. Inst. Elect. Engrs*, **91/1**, 340.
 Takagi, M. 1939 *Sci. Rep. Tōhoku Univ.* **28**, 26.
 Taylor, G. I. 1934 *Proc. Roy. Soc. A*, **145**, 362.
 Webster, W. L. 1925 *Proc. Roy. Soc. A*, **107**, 496.

Electron traps and dielectric changes in phosphorescent solids

By G. F. J. GARLICK AND A. F. GIBSON

Department of Physics, University of Birmingham

(Communicated by M. L. Oliphant, F.R.S.—Received 2 April 1946)

Some luminescent solids, such as zinc sulphide with a specific impurity, exhibit considerable increases in their dielectric constant and loss when excited in the presence of a small electric field by ultra-violet light or other suitable radiation. The effect is found to be confined to those phosphors which show photoconductivity associated with the luminescence process. The total change in dielectric constant of such phosphors may increase by as much as 75 % of the normal value (about 8) for quite moderate intensities of excitation. The loss increase, expressed as the change in equivalent a.c. conductance, may change from less than 10^{-8} to 10^{-5} ohm $^{-1}$ during excitation.

Experimental investigations and theoretical considerations show that the dielectric changes occur when electron traps are filled by electrons. The relatively low binding energy of the trapped electron to its trap (from 0.1 to 0.7 eV in most zinc sulphide phosphors) permits large displacements of the electron from its mean position when an external field is applied. This gives rise to the increase in dielectric constant. Increase of dielectric loss of the phosphor is associated with the existence of a finite relaxation time of the filled electron traps which causes phase differences between the variation of the applied alternating field and the movement of the trapped electrons. This gives rise to energy absorption from the applied field by the filled traps. It is found from experiment that electrons moving in the conduction energy levels of the phosphor crystal make a negligible contribution to the dielectric effects when compared with the contribution from trapped electrons.

The dielectric changes are found to increase rapidly with temperature following an exponential variation. They are also dependent on the applied field frequency over the frequency range comparable in order to the reciprocal of the relaxation time of filled electron traps which is found to be of the order of 10^7 sec. $^{-1}$. The relaxation time does not vary rapidly with the temperature. The dielectric changes show small variation with the wave-length of the exciting light between 0.35 and 0.41μ for most zinc sulphide phosphors, but decrease rapidly as the wave-length decreases to 0.25μ , due to a rapid rise in the absorption coefficient of the phosphors. The long wave-length limit for production of the dielectric changes corresponds to the wave-length limit for photoconductivity and for the filling of electron traps and is dependent upon the constitution and structure of the particular specimen.

The electron-trap theory of phosphorescence established by previous workers enables many of the dielectric effects to be correlated with the luminescence processes in phosphors. The studies confirm the results of previous workers and enable their observations on the dielectric changes to be given a theoretical interpretation.

GENERAL INTRODUCTION

It has been known since the early work of Lenard & Saeland (1909) that some luminescent materials, such as zinc sulphide with an impurity activator, show an increase in dielectric constant in an applied electric field when excited to emit luminescence. The effect has been studied by other workers (Gudden & Pohl 1920; Herweg 1923; Wilde 1923; Schmidt 1923; Goos 1939; and others); the corresponding change in dielectric loss has been studied by Gisolf (1939). From the results of these investigations it was found that the effects were limited to those phosphors in which photoconductivity occurred in association with the luminescence process. The increase in dielectric constant was observed to be of the same order as the normal value (about 8) in the case of some zinc sulphide specimens, this increase being for quite normal intensities of excitation.

The present studies confirm the previous findings and provide in addition new experimental facts which enable a theoretical model to be established to explain the origin of the dielectric changes in phosphors. Previous workers in this laboratory have shown that the phosphorescence processes in photoconducting phosphors such as zinc sulphide and the alkaline earth sulphides depend on the storage action of thermally metastable centres in the phosphors known as electron traps. These traps capture electrons liberated from the luminescence centres of the phosphor by the excitation and retain them until thermal energy added to the traps by local lattice vibrations enables the electrons to escape and give rise to phosphorescence by recombining with luminescence centres. The mean time spent by an electron in a trap determines the time at which it gives rise to phosphorescence (Randall & Wilkins 1945). It has also been shown that this electron-trap mechanism can govern the phosphorescence characteristics at very small decay times of the order of 1 msec. (Garlick & Wilkins 1945). By the use of theoretical ideas formulated by these workers it has been possible to interpret the dielectric changes in phosphors in terms of the electron-trap processes and to relate the dielectric phenomena to the luminescence properties of phosphors. Phosphorescence which is governed by the action of electron traps is markedly dependent on the conditions of excitation and the phosphor temperature. A similar dependence is shown by the dielectric changes during fluorescence, phosphorescence and thermoluminescence of the material. From the correspondence between the luminescence and dielectric effects it is inferred that the dielectric changes arise from the behaviour of electrons during the time in which they are separated from the luminescence centres. This means that the dielectric changes may be due to the properties of electrons moving through the conduction band of the phosphor or to the effects of trapped electrons.

THEORETICAL INTRODUCTION

As the theoretical ideas developed in these studies include many of those established by previous workers (Randall & Wilkins 1945), a brief survey of the electron-trap theory of phosphorescence is given embodying portions relevant to the interpretation of the dielectric phenomena. In luminescent solids activated by the inclusion of a small amount of a specific impurity, the seat of the luminescence emission is in the impurity atoms or ions which are dispersed throughout the matrix lattice of the solid. The heat treatment of the phosphor in preparation effects this dispersal and also gives rise to potential defects in the periodic field of the crystal lattice (Mott & Gurney 1940). Some of these defects will be capable of trapping and storing electrons liberated from the luminescence centres by the exciting light. The exact nature of these trapping centres is not known, but there is some evidence that they are associated with the neighbourhood of the impurity centres; it is hoped to publish such facts shortly. The number of effective traps is estimated to be of the same order as the number of luminescence centres. Using the collective electron model for insulating solids now generally accepted, the luminescence mechanism

is described as follows: excitation of the phosphor frees electrons from the luminescence centres, and these electrons move through the conduction band of the crystal either returning to luminescence centres or being trapped in the electron traps. The energy states of the traps are assumed to lie just below the lowest states of the conduction band. In most phosphor specimens there are many trapping energy states of various depths below the conduction band.

The time spent by an electron in a trap will depend on the depth of the trap below the conduction band and on the temperature. If p is the probability per unit time that an electron will escape from a trap of depth E at a temperature T and t is the mean time spent in the trap then p and t are given by

$$p = 1/t = s \exp(-E/kT), \quad (1)$$

where s is a constant which may, however, vary slowly with temperature and k is Boltzmann's constant. The constant s has the dimensions of a frequency and is often considered as bearing a simple relation to the natural frequency of vibration of the filled electron trap. From phosphorescence and thermoluminescence experiments it is estimated to be of the order of $10^{8 \pm 1} \text{ sec.}^{-1}$ for zinc sulphide phosphors. It is associated later with the relaxation time of the filled electron traps as measured by dielectric experiments. A method of finding the number of electron traps at different depths below the conduction band has been developed by Randall & Wilkins which involves the measurement of the variation of thermoluminescence glow emission with temperature when a phosphor is warmed at a uniform rate in the dark after excitation at a low temperature so that nearly all traps are saturated with electrons. The luminescence emission intensity of the phosphor is proportional to the rate of escape of electrons from traps, which is proportional to the number of trapped electrons and their probability of escape p . Thus if there are n trapped electrons in traps of depth E the emission intensity I is given by

$$I = -dn/dt = pn = ns \exp(-E/kT). \quad (2)$$

This set of relations is based on the assumption that there is no retrapping of escaping electrons. For practical cases where there are many traps of different depths in the phosphor the equations must be integrated over all values of E and n . The number of electrons trapped will depend on the excitation conditions, the temperature and the form of the trap distribution with respect to depth. It can be shown from simple considerations of equilibrium conditions that for steady excitation at a fixed temperature the number of electrons in traps of depth E is given by

$$n = N_E / \{1 + s/A \exp(-E/kT)\}, \quad (3)$$

where N_E is the total number of traps of depth E and A is a constant proportional to the exciting light intensity. When there are many traps of different depths the total number of trapped electrons is given by the integral of n over all E values:

$$N = \int_0^{E_{\max.}} n dE = \int_0^{E_{\max.}} N_E / \{1 + s/A \exp(-E/kT)\} dE, \quad (4)$$

where N is the total number of trapped electrons. Thus by means of equations (2) and (3) the decay of phosphorescence with time may be determined for different excitation intensities. N_E for each trap depth can be found from the thermoluminescence experiment. Randall & Wilkins have given the derived phosphorescence decay equations for saturated conditions (that is $n = N_E$), and for long decay times only for certain variations of N_E with E . One of these relations is relevant to the study of dielectric changes in phosphors during phosphorescence. It is that obtained for N_E varying exponentially with E between certain values of E . If, therefore, $N_E \text{ const. exp}(-\alpha E)$, where α is a constant, then the intensity of phosphorescence at a decay time t is given by

$$I = \text{const.}/t^{\alpha kT+1}. \quad (5)$$

The constant α can be obtained from the thermoluminescence experiment and can be compared with the value obtained from measurement of phosphorescence decay at a fixed temperature. Correlations between the two derivations of the constant have been given for several different phosphors by Randall & Wilkins (1945, p. 398).

THEORETICAL INTERPRETATION OF DIELECTRIC CHANGES IN PHOSPHORS

The observed increases in dielectric constant and dielectric loss of the phosphors studied are found to occur only when luminescence is accompanied by photoconduction processes. From this it is inferred that the effects are due to the action of electrons which have been freed from the luminescence centres by the excitation and are either moving in the conduction band or are captured in electron traps. Electrons moving in the conduction band are known to give rise to conduction processes in the phosphor in the presence of an applied field and might therefore contribute to the dielectric changes observed when the field is alternating. The motion of such electrons will give rise to an increase in dielectric loss and also to increase in dielectric constant if they are subject to suitable constraining forces so that they do not move in phase with the field. Dielectric changes might also occur when an electron is captured in a trap, since the binding energy associated with its captured state is usually small in phosphors (from 0.1 to 0.8 eV; see Randall & Wilkins 1945). Thus a trapped electron can be displaced from its mean trapped position by the action of an external applied field. This displacement might be referred to as the polarization of a filled electron trap. It would cause increase in the dielectric constant of the phosphor and also increase in dielectric loss if the displacement motion was not in phase with the applied field variation. Such a phase difference would arise if the natural vibration frequency of the filled trap system was of the same order as the applied field frequency.

Experimental evidence indicates that the dielectric changes are predominantly due to the displacements of trapped electrons. Some of the most significant experimental facts are given here in summarized form to emphasize the negligible effect of electrons in the conduction band on the dielectric properties of phosphors compared with the effect of filled electron traps.

(a) When the phosphor is excited at a fixed temperature to an equilibrium state the total fluorescence emission is found to increase linearly with the intensity of excitation. At normal intensities of excitation the fluorescence is mainly due to electrons which have not been captured by electron traps but which return directly to luminescence centres after excitation. Although the fluorescence emission shows this simple linear variation with intensity of excitation and therefore the increase in the number of electrons in the conduction band also varies regularly with excitation intensity, yet the dielectric changes show a saturation effect at quite low intensities. This is shown by the results given in figure 3. The same saturation effect occurs for the number of trapped electrons and for the phosphorescence after cessation of excitation. The relation between the dielectric changes and the excitation intensity is found to be identical with that for the variation of the number of trapped electrons with intensity of excitation.

(b) When a phosphor is excited at liquid-air temperature most of the electron traps are filled and will remain filled after excitation has ceased if the low temperature is maintained. After excitation ceases the phosphorescence decays rapidly to a negligible value, but the dielectric changes decay to a steady value which is still about 70–80 % of the value during excitation. Such a large residual change in dielectric constant or loss cannot be due to electrons present in the conduction band, as there is no phosphorescence emission due to their capture by luminescence centres. If after a considerable time the phosphor is irradiated with infra-red light, then a burst of luminescence occurs due to the release of electrons from traps. The dielectric changes decrease to zero with no initial enhancement such as would be expected if they were due to conduction band electrons. This behaviour has been observed by other workers (Goos 1939).

(c) From measurements of the frequency dependence of the dielectric changes, values of the relaxation times of the polarizable systems causing the dielectric changes are obtained and are of the order of 10^{-7} sec. Such large values are not likely to be associated with electrons in the conduction band, but it is not difficult to associate them with filled electron traps as previous workers have given estimates of the values of the natural vibration frequencies of filled traps as being $10^{6\pm 1}$ sec.⁻¹ (Randall & Wilkins 1945, p. 383). This frequency will be closely related to the reciprocal of the relaxation times of the filled traps.

From this and other experimental evidence described later it is concluded that the dielectric changes are due to the properties of trapped electrons and that electrons in the conduction band make no significant contribution. The increase in dielectric constant therefore involves the displacement of trapped electrons from their mean positions. An estimate of the magnitude of this displacement can be obtained from the number of filled traps and the percentage increase of the dielectric constant. The number of trapped electrons is certainly less than 10^{18} per c.c. (Randall & Wilkins 1945, p. 404), and the dielectric constant can be more than 75 % of its normal value. Thus the electron displacement to effect such a change must be several lattice spacings of the phosphor in magnitude. In order to explain the corresponding

increase in dielectric loss the trapped electrons must be capable of absorbing small quanta of energy from the applied field which implies the existence of a number of energy states for the trapped electron which are sufficiently close together to allow the energy absorption. Such a configuration implies a large effective trap diameter. Large orbits associated with trapped electrons have been postulated by previous workers to explain some of the properties of semiconductors (Mott & Gurney 1940, p. 166). The magnitude of the trapped electron displacement will depend on the depth of the electron trap as will its capacity to absorb energy from the applied field. The dielectric changes would be expected to decrease with increase of trap depth together with a similar decrease in the relaxation time of the filled trap.

Since the binding conditions of the trapped electron are affected by the temperature (e.g. equation (1)), it would be expected that the dielectric changes would depend upon the temperature of the phosphor. It is found from the experimental data that the dielectric changes increase exponentially with the temperature above about 200° K., but below this point they do not vary rapidly with temperature. These facts might be used to predict the form of the potential-energy functions of trapped electrons, but considerable extra data are required for this purpose and the functions have not been estimated as yet. The above theoretical discussions show that if the dielectric changes in phosphors are due to filled electron traps then the magnitude of the effects will depend on the number of electrons trapped, the trap-depth distribution, the relaxation times of the filled traps, the applied field frequency and particularly the temperature of the phosphor. That is, the filled electron traps can be considered as polarizable and absorbing systems in the matrix phosphor material, and that they consist of a loosely bound electron which is readily displaced by applied fields and an ionic or atomic system of much greater mass or inertia to which the electron is bound.

EXPERIMENTAL METHOD

For measurements of dielectric changes or luminescence the phosphor is mounted in a very thin layer on a rhodium-plated copper surface forming part of a small German silver Dewar vessel. This apparatus has been used for previous work (Garlick & Wilkins 1945) and is particularly suited to the variation of the phosphor temperature over the wide range from 90 to 600° K. The temperature variation is effected by filling the inner tube of the Dewar vessel with liquid air or by the use of a small electric heater attached to the rear of the surface on which the phosphor is mounted. The phosphor, which is *in vacuo*, is excited through a quartz window in the outer structure of the Dewar vessel. For dielectric measurements the phosphor is covered by a thin glass slip with a graphite coating in grid form on its outer surface through which the exciting radiation can pass. The phosphor is thus enclosed in a condenser cell formed by the rhodium-plated surface and the graphite grid and forms part of the dielectric of the cell the rest being due to the glass of the cover-slip. Electrical connexion is made to the graphite and to the plated-copper surface, the latter being

maintained at earth potential. A thermocouple is attached to the front of the copper surface to measure the temperature of the phosphor in the cell. The source of exciting radiation in most of the experiments is a high-pressure mercury arc lamp of the Osira type used in conjunction with suitable filters for the wave-length desired. In most of the experiments a wave-length of 0.365μ has been used.

Apparatus for the measurement of dielectric changes

As the dielectric changes of the phosphor had to be measured while varying with time or temperature the conventional methods of determining dielectric constant or loss were not suitable and electrical circuits were devised which would measure, independently of each other, the dielectric constant or loss changes with time or temperature. A diagram of the electrical circuits of the measuring apparatus is given in figure 1. A conventional tuned anode-tuned grid radio-frequency oscillator

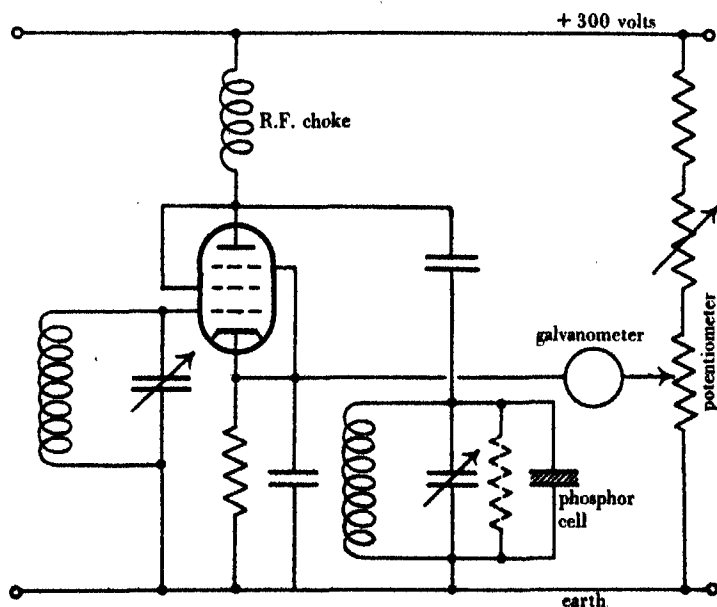


FIGURE 1. Circuit diagram of dielectric change-measuring apparatus.

is loosely coupled to the grid circuit of the valve shown in figure 1. This valve is a pentode (Mazda SP 41) working as a triode and as an anode bend detector; its normal grid bias provided by the cathode load is about -13 V. The anode circuit of the valve contains a tuned circuit in which the phosphor cell of the Dewar apparatus is connected in parallel with a variable condenser and suitable inductance. The anode voltage is supplied through a radio-frequency choke from stabilized power supplies which also feed the master oscillator. By means of stabilization and careful screening random fluctuations during measurement can be reduced to a very small minimum. The circuit operates as follows: A radio-frequency voltage due to the oscillator is

established in the tuned grid circuit of the valve and results in the development of a radio-frequency voltage across the tuned anode circuit. When the grid and anode circuits are in tune a sharp resonance state occurs due to feed-back through the valve. The resonance curve of radio-frequency voltage across the tuned anode circuit for changes in capacity of the circuit is given schematically in figure 2. Changes in dielectric constant in the circuit cause a capacity change in the circuit, while changes in dielectric loss cause a flattening of the resonance curve. The voltage changes across the anode circuit due to dielectric changes are recorded by the galvanometer in the cathode circuit of the valve. A potentiometer system is provided to annul the normal cathode voltage existing when no changes occur.

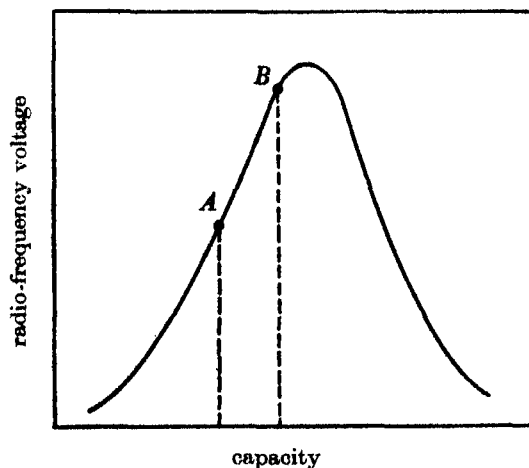


FIGURE 2. Resonance curve for tuned anode circuit of figure 1.

In order to measure capacity changes, which are due to dielectric constant changes, separately from the effects due to dielectric loss the anode circuit is tuned by means of its variable condenser to a point on the resonance curve shown by *A* in figure 2, and a relatively low resistance is connected in parallel with the phosphor cell. Its value is chosen to be much lower than the minimum a.c. resistance of the cell due to dielectric loss increase during luminescence. Although this procedure lowers the sensitivity of the system to capacity changes it is still sufficient for practical measurement and enables the effect of loss changes during capacity measurement to be reduced to a minimum as a contribution to the galvanometer deflexion. This contribution is found to be less than 1 % in experiment. The linearity of the galvanometer response to capacity changes is tested by the use of standard condensers across the tuned anode circuit, the same condensers being used for calibration of the capacity changes in micro-microfarads. The phosphor cell has a capacity of about $10\ \mu\mu\text{F.}$ in the dark; the apparatus can detect a change in capacity of $0.01\ \mu\mu\text{F.}$

For the measurement of dielectric loss changes the parallel resistance inserted for capacity change measurements is removed. The circuit is then tuned to a point *B*

shown in figure 2 near to the peak of the resonance curve. Under these conditions the change in voltage across the tuned circuit recorded by the galvanometer will be mainly due to the flattening of the resonance curve when loss changes occur. The contribution from changes in capacity is found to be less than 2 % of the recorded change. The calibration of the galvanometer for loss measurements is effected by inserting standard non-inductive resistances across the tuned circuit. There is some departure from linearity between loss change and galvanometer response but this is included in calibration.

Measurements of dielectric changes at different frequencies of the applied field

The apparatus described above has been used mainly at a frequency of 3 Mcyc./sec. In addition, measurements of dielectric changes have been made at a lower frequency of 1 Mcyc./sec., using the same apparatus. In general, however, it is not possible to extend the use of the apparatus to all desired frequencies for complete measurement of the dielectric changes during fluorescence, phosphorescence and thermoluminescence. Measurements of the dielectric changes have been made for steady excitation of fluorescence at fixed temperatures over a frequency range from 100 kcyc./sec. to 10 Mcyc./sec. This range proves sufficient to include the anomalous frequency variation of the dielectric changes due to the particular values of relaxation times of the polarizable systems. For these measurements at different frequencies of the applied field the phosphor cell is connected to the circuit of a conventional 'Circuit magnification or Q meter', by means of which the dielectric constant and loss changes are recorded.

All changes in dielectric constant are denoted by $\Delta\epsilon$ in the description of experimental results. Loss changes are measured as the change in equivalent parallel conductance $\Delta\gamma$ of the cell in units of reciprocal ohms.

Luminescence measurements

For the determination of the luminescence characteristics and electron-trap distributions of the various phosphor specimens examined the methods described previously have been employed (Randall & Wilkins 1945).

EXPERIMENTAL RESULTS

Studies of the dielectric changes during luminescence in many different specimens of zinc sulphide and zinc-cadmium sulphides have been made. The results given here are for a few of these phosphors only but are typical of those obtained for all specimens. In all cases the dielectric characteristics of the phosphor cell in the dark, unexcited state are determined so that the changes due to the luminescence processes only can be deduced from the results. Such control experiments are made over the range of temperature from 90 to 600°K. For each specimen the fluorescence-temperature, the saturated phosphorescence-time and the thermoluminescence-temperature characteristics are measured in order to define the luminescence characteristics of each specimen. In all cases it is found that the dielectric constant

and loss of the unexcited phosphor show no marked temperature effects over the range investigated and no anomalous variation with the applied field frequency. Above 600° a temperature is reached at which loss changes due to semi-conduction processes in the phosphor occur.

Absolute measurements of the dielectric changes are difficult to obtain, as all the phosphors are in powder form and mounted in a very thin layer so that the accuracy of estimating cell dimensions and effective volume of the dielectric due to the phosphor is relatively low. Thin layers are essential to obtain complete and uniform excitation of the specimen thus avoiding anomalies due to non-uniform excitation (Goos 1939). Measurements made show that the normal dielectric constant of phosphors of the zinc sulphide type is about 8 and that for high excitation the value increases to about 14 for some of the specimens. Previous workers have claimed to have observed increases of the order of 100 % of the normal value (Gudden & Pohl 1920). In view of the above difficulties the measurements of the dielectric changes, with the exception of those recorded in figure 3, have been made in arbitrary units. Relative accuracies of measurement are high, of the order of 1 or 2 %. Dielectric loss changes can be measured in absolute units as previously stated, but the low values for the unexcited phosphor are inaccurate because of their small magnitude. The values are certainly less than 10^{-8} ohm $^{-1}$.

(1) *Variation of the dielectric changes with the intensity of excitation at fixed temperature*

When a phosphor is excited for a sufficiently long time at a fixed temperature, in this case room temperature, by a steady source of ultra-violet light it reaches a state of equilibrium when its emission intensity is constant and is proportional to the excitation intensity. There are accompanying changes in dielectric constant and loss, but these changes tend to saturate at quite small excitation intensities. A typical variation of the dielectric constant change with excitation intensity is given in figure 3 for a zinc sulphide phosphor activated by silver impurity. The ordinate of the figure is the change in dielectric constant $\Delta\epsilon$ in absolute units, the scale being accurate to ± 10 % in absolute value. The abscissa is the exciting light intensity in arbitrary units. Measurements such as those given in figure 3 have been made for all specimens over intensity ranges of more than 1000:1. From the experimental results of figure 3 it can be seen that saturation effects occur at low intensities and that the curve flattens out more as the intensity increases. It is found that at high intensities the dielectric changes $\Delta\epsilon$ and $\Delta\gamma$ vary linearly with the logarithm of the exciting light intensity I . From the experimental results a more complete empirical formula relating the dielectric changes and the excitation intensity has been derived. It is

$$\Delta\epsilon \text{ or } \Delta\gamma \text{ is proportional to } \log(Ib + 1), \quad (4)$$

where b is a constant which can be determined from the results at high intensities. It is obtained by plotting $\Delta\epsilon$ or $\Delta\gamma$ against $\log I$ and producing the straight line for the high intensities when its intersection with the abscissa for $\Delta\epsilon$ or $\Delta\gamma$ equal to

zero is at $-\log(b)$. The product Ib in equation (6) is denoted by A' , this factor still being proportional to the excitation intensity. Experimental results are given in figures 4 and 5 to show the validity of the empirical relation for six typical phosphors. The ordinates of the figures 4 and 5 are the change in dielectric constant $\Delta\epsilon$ and the change in dielectric loss $\Delta\gamma$ respectively. The abscissa of each figure is the quantity $\log(A' + 1)$, the value of A' being determined for each phosphor as described above. It is found that the number of filled electron traps, as measured by the thermoluminescence experiment, varies with the excitation intensity according to the relation given by equation (6) and the value of A' obtained from the thermoluminescence measurements agrees with the values obtained for the variation of the dielectric changes with excitation intensity.

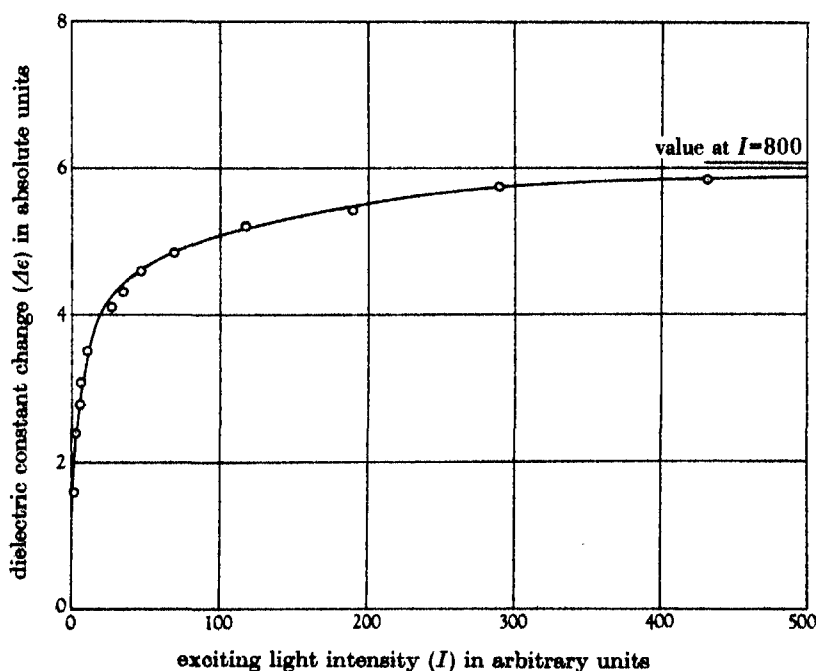


FIGURE 3. Variation of dielectric constant change with exciting light intensity for atypical phosphor (ZnS-Ag) at 291° K. (Absolute accuracy of ordinate units is $\pm 10\%$.)

The above experimental results indicate that the dielectric changes are proportional to the number of filled electron traps. Using the conceptions developed in the theoretical section above a derivation of equation (6) can be obtained. If the temperature, and the field frequency, remain fixed and the variation of the dielectric changes with trap depth is neglected, then the dielectric changes at any intensity will be proportional to the number of filled traps. This number will be given by equation (4). If N_E is constant or is varying slowly with E then the equation has the solution

$$N = N_E kT \log [(Ax_m + 1)/(A/s + 1)],$$

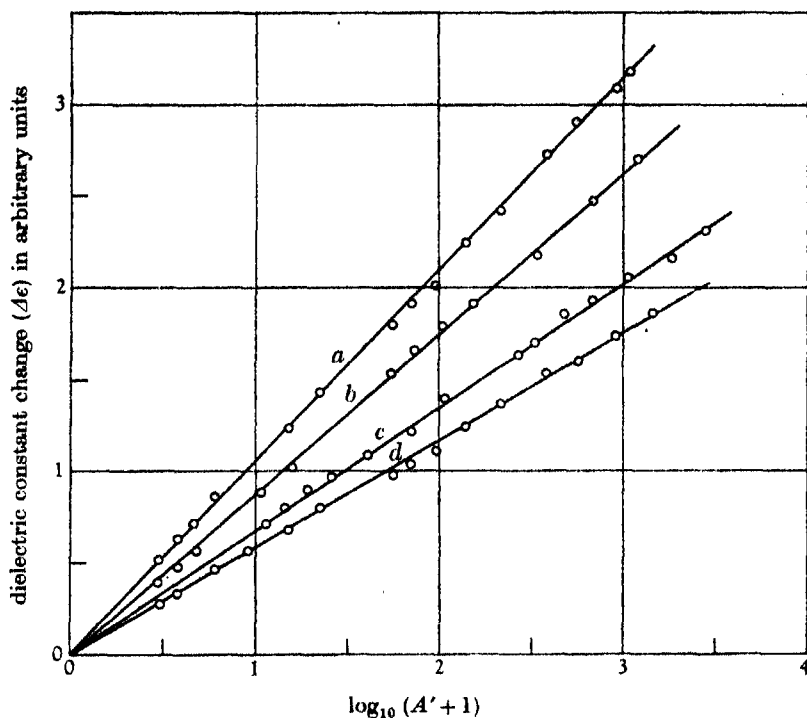


FIGURE 4. Variation of dielectric constant change with exciting light intensity for various phosphors at room temperature (291° K). (a) ZnS-CdS-Cu (short after-glow); (b) ZnS-Cu (long after-glow); (c) ZnS-Ag (short after-glow); (d) ZnS-Zn (short after-glow). A' = constant factor (b) \times exciting light intensity (I).

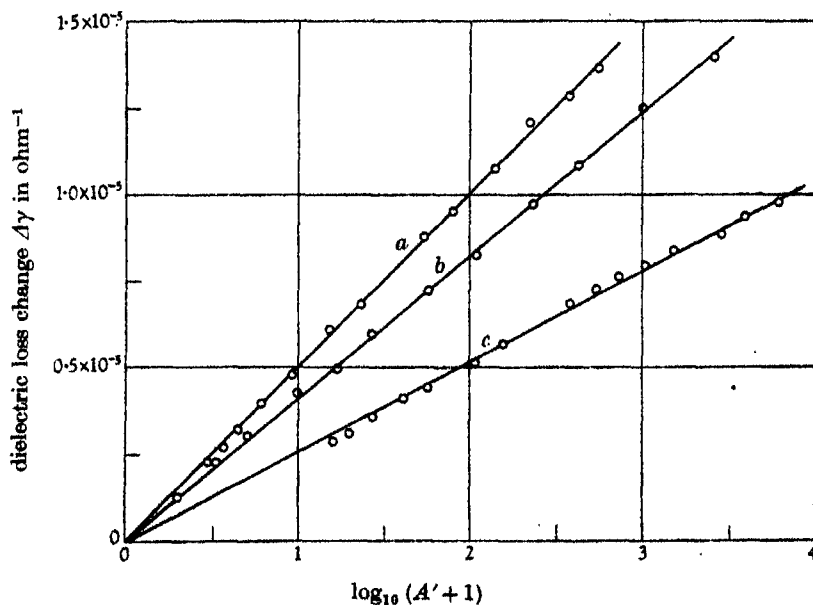


FIGURE 5. Variation of dielectric loss change with exciting light intensity for various phosphors at room temperature (291° K). (a) ZnS-Cu (long after-glow); (b) ZnS-Ag (short after-glow); (c) ZnS-CdS-Cu (long after-glow). A' = constant factor (b) \times exciting light intensity (I in arbitrary units).

where $x_m = s^{-1} \exp(E_{\max}/kT)$. From experiment it is found that over the intensity range used A is much smaller than s so that the above expression becomes

$$N = N_E kT \log(A' + 1), \quad (7)$$

where $A' = Ax_m$. As the dielectric changes $\Delta\epsilon$ and $\Delta\gamma$ will be proportional to N this derivation is equivalent to the relation found empirically in equation (6). The agreement shows that the assumption that the variation of dielectric changes with the trap depth is slow is justified. The assumption that N_E is constant or varies but slowly with E is also justified if all traps are considered.

(2) *Decay of the dielectric changes with the decay of phosphorescence*

If the phosphor specimen is excited to an equilibrium state at room temperature and then the excitation is cut off the phosphorescence will decay in intensity with time. The intensity-time relation will be determined by the extent to which the traps are filled by the excitation. The dielectric changes also decay with time during phosphorescence. To determine the relation between the dielectric changes and the decay of phosphorescence the following experiments were carried out:

- (i) Measurement of the decay of phosphorescence with time.
- (ii) Measurement of the decay of the dielectric constant and loss changes with time.
- (iii) Measurement of the decrease in the number of trapped electrons with time by means of the thermoluminescence experiment.

Experimental results for two of the phosphors studied are given in figures 6 and 7. The co-ordinates are logarithmic, the abscissa of each figure being the decay time. The ordinates of each graph give the intensity of phosphorescence, the dielectric changes $\Delta\epsilon$ and $\Delta\gamma$ and the number of trapped electrons, all quantities being expressed in arbitrary units. From an inspection of the curves of these two figures it is evident that the slopes of the dielectric constant and loss change curves are about the same as the slopes of the curves giving the decay in the number of trapped electrons with time. However, these several slopes are different from that of each respective phosphorescence curve by about unity. The results show that the dielectric changes depend mainly upon the number of filled electron traps.

The difference in slope between the curves for the dielectric changes and number of trapped electrons and the curve for the decay of phosphorescence may be explained by reference to equation (5) above. The number of electrons in traps at a decay time t when the phosphorescence intensity is I will be given by

$$N = \int_t^\infty I dt.$$

The phosphorescence decay of the two phosphors in question follows the relation of equation (5) and therefore the integral becomes

$$N = \int_t^\infty (\text{const.}/t^{akT+1}) dt = \text{const.}/t^{akT}.$$

Thus the respective slopes should differ by unity when the decay processes are presented as in figures 6 and 7 on logarithmic co-ordinates. When considering the variation of the dielectric changes with excitation intensity it was assumed that N_E varied slowly with E or was constant, whereas in the present instance an exponential variation of N_E is assumed. In the interpretation of the results given in §(1) above it is justified to consider N as essentially constant, as all traps of all depths are involved; in the case of phosphorescence decay at long decay times only deep traps are involved, and their number may vary exponentially with E without invalidating the assumptions of §(1) for all traps.

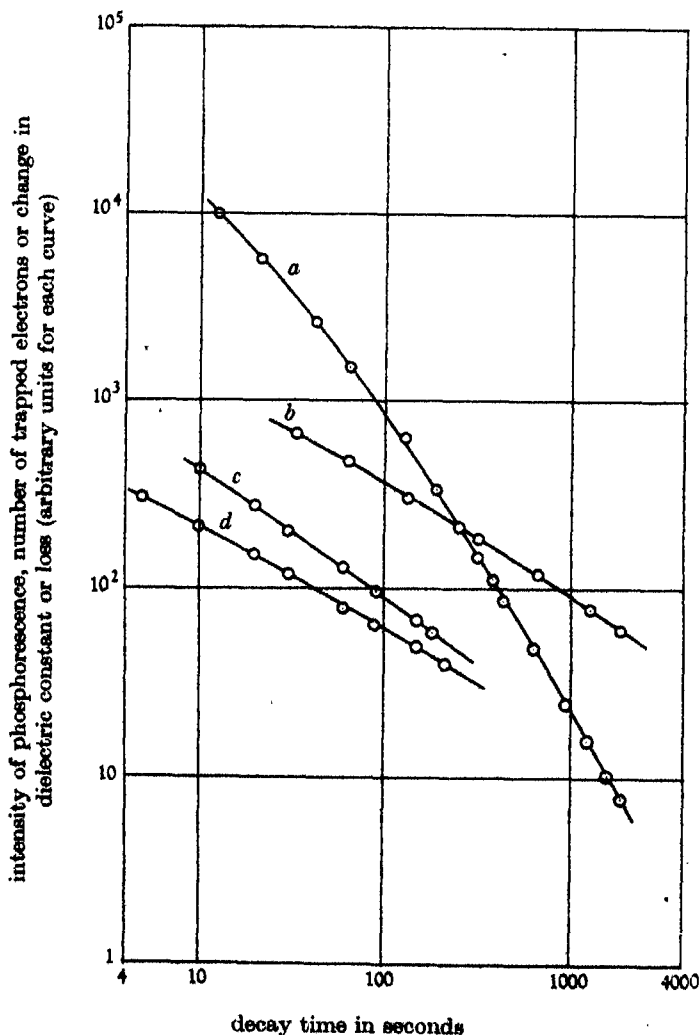


FIGURE 6. Decay of dielectric changes, phosphorescence and number of trapped electrons with time for ZnS-Cu phosphors at 291° K. (a) decay of phosphorescence; (b) decay of number of trapped electrons; (c) decay of dielectric loss change; (d) decay of dielectric constant change.

The approximate nature of the agreement between the slopes of the curves for the dielectric changes and those for the decay of the number of trapped electrons indicates that the effect of trap depth on the dielectric changes may not be negligible. The dielectric changes will decrease with the depth of the filled traps, and as the decay proceeds the electrons remain in the deeper traps so that the dielectric changes have a more rapid decay than that which would occur if there were no effect due to trap depth.

In addition to the measurement of phosphorescence decay and the corresponding decay of the dielectric changes some qualitative experiments have been made at

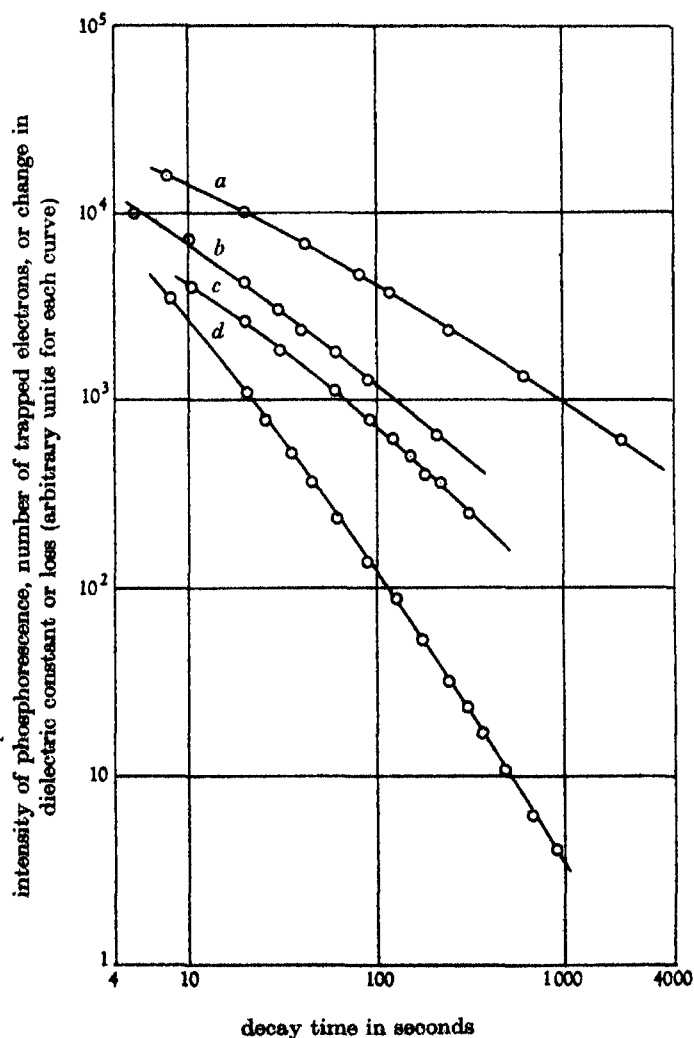


FIGURE 7. Decay of dielectric changes, phosphorescence and number of trapped electrons with time for a ZnS-CdS-Cu phosphor at 291° K. (a) Decay of number of trapped electrons; (b) decay of dielectric constant change; (c) decay of dielectric loss change; (d) decay of phosphorescence.

liquid-air temperature which have already been briefly described in the theoretical section above. They involve the release of trapped electrons by infra-red light. Similar experiments made at room temperature confirm earlier work (Goos 1939; and others). The irradiation of a phosphor during phosphorescence decay with infra-red light can cause a temporary enhancing of the phosphorescence with a subsequent decay which is more rapid than its normal progress. In some cases the initial enhancing does not occur but the decay becomes more rapid. The occurrence of one or both of these phenomena depends on the particular specimen used. However, in no case is any initial enhancing of the dielectric changes observed. The enhancing of phosphorescence may also be produced by a strong electric field which can release electrons from traps. In this case an earlier worker has shown that no initial increase in dielectric constant ever occurs (Herweg 1923). These experiments show that the dielectric effects are due to trapped electrons and that electrons in the conduction band contribute a negligible amount to the effects.

(3) *Variation of the dielectric changes during thermoluminescence emission*

If the phosphor specimen, mounted in the Dewar apparatus, is cooled to liquid-air temperature, excited and then warmed at a uniform rate in the dark (2.5°K/sec. in all experiments) it emits thermoluminescence which varies in a characteristic way with the temperature for each specimen. The variation gives a representation of the number of electron traps of different depths in the phosphor if excitation at the low temperature is high enough to saturate the traps. During the emission of thermoluminescence changes also occur in the dielectric constant and loss of the specimen. In the experiments described below these changes during thermoluminescence have been measured. In addition, the temperature variation of the conduction current in the phosphor when a static applied field is present during thermoluminescence has been measured as representing the number of electrons moving through the conduction band at any instant during the thermoluminescence emission. The variation of the dielectric changes during thermoluminescence together with the emission and the variation of the d.c. conduction current are given in figure 8 for a typical phosphor. The abscissa of this figure is the temperature. The ordinate represents for curve (a) the intensity of thermoluminescence, for curve (b) the change in dielectric loss $\Delta\gamma$, for curve (c) the change in dielectric constant $\Delta\epsilon$, for curve (d) the magnitude of the conduction current for a d.c. applied field. It can be seen that there is good agreement between curves (a) and (d) but that there is no simple relation between curves (b) and (c) for the dielectric changes and curve (a) for the variation of thermoluminescence with temperature. The dielectric changes do, however, reach zero at the same temperature as the thermoluminescence and the conduction current showing that there are no changes when the electron traps are all empty.

To interpret the curves (b) and (c) of figure 8 it is assumed that the dielectric changes depend on the number of filled traps, but in addition it is assumed that there

may be a variation of the dielectric changes with temperature. To ascertain the nature of the temperature variation the first assumption is used. At any temperature T_G during the thermoluminescence emission the area remaining under the curve (a) between T_G and the temperature at which emission ceases is proportional to the number of trapped electrons. In the absence of a temperature variation of the dielectric changes then the ordinate of either curve (b) or (c) at T_G would be proportional to this area on the above assumption. Thus the temperature function for the dielectric changes may be found by dividing the ordinate of curves (b) or (c) by

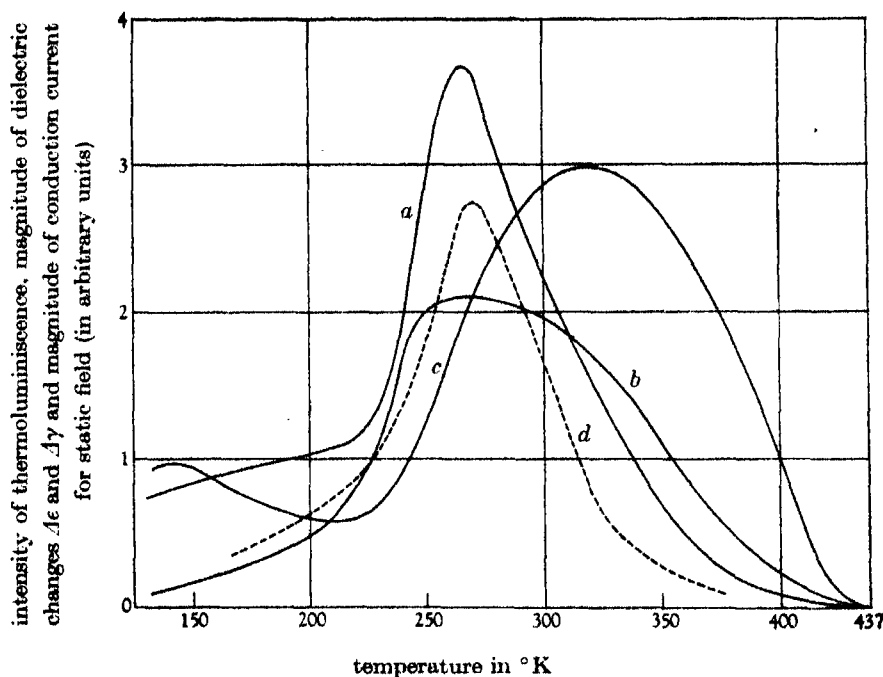


FIGURE 8. Variation of thermoluminescence emission and corresponding dielectric changes with temperature for ZnS-CdS-Cu phosphor after excitation at 90° K. (a) Thermoluminescence-temperature curve; (b) dielectric loss change-temperature curve; (c) dielectric constant-temperature curve; (d) current flowing in phosphor during thermoluminescence in static applied field.

the area left under the thermoluminescence curve. If this is carried out at all temperatures the intrinsic variation of the dielectric changes with temperature may be obtained. The temperature functions so derived are given for the dielectric constant and dielectric loss changes in figures 9 and 10 respectively. Figure 9 has temperature as the abscissa and the temperature function of the dielectric constant change as the ordinate, denoted by $f_{\Delta\epsilon}(T)$. The ordinate is given with linear and logarithmic scales corresponding to curves (b) and (a) respectively in order to show the form of the function. The abscissa of figure 10 is the inverse of the temperature and its ordinate the temperature function for the dielectric loss change on

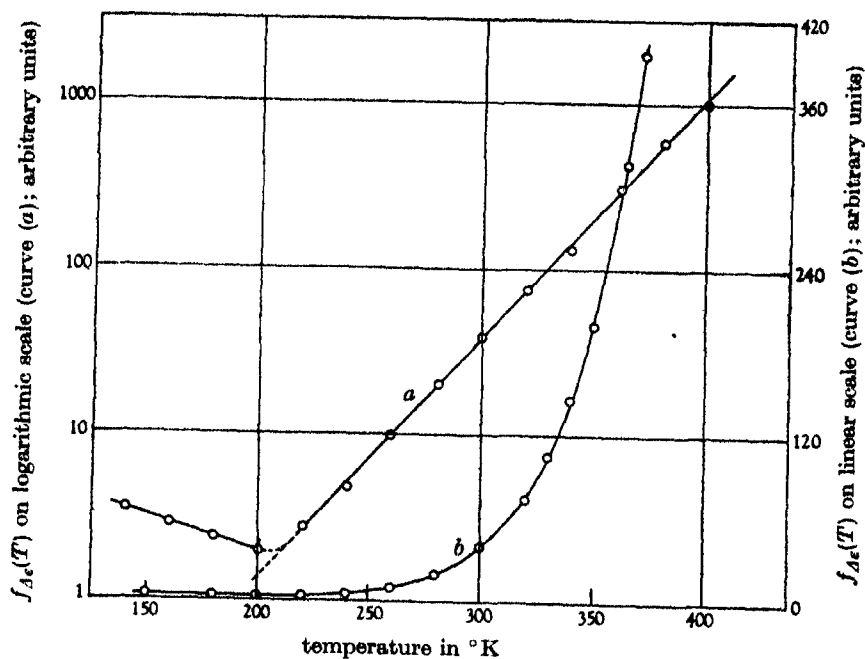


FIGURE 9. Temperature function $f_{\Delta\epsilon}(T)$ for dielectric constant change with temperature for ZnS-CdS-Cu phosphor. ($f_{\Delta\epsilon}(T)$ is derived from figure 8.)

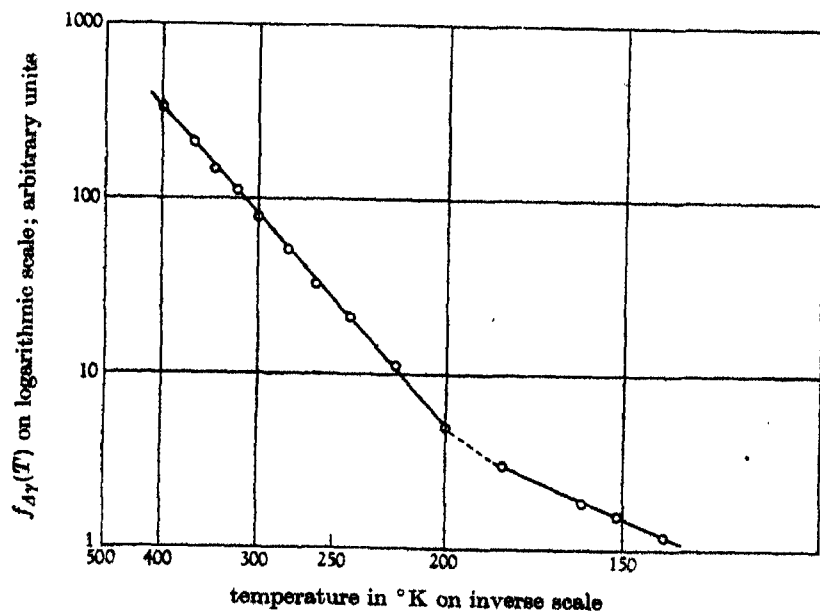


FIGURE 10. Temperature function $f_{\Delta\gamma}(T)$ of dielectric loss change with temperature for ZnS-CdS-Cu phosphor. ($f_{\Delta\gamma}(T)$ is derived from figure 8.)

a logarithmic scale, the function being denoted by $f_{\Delta\epsilon}(T)$. From the curves of figures 9 and 10 the respective temperature functions are found to be

$$\begin{aligned} f_{\Delta\epsilon}(T) &= \text{const. exp } (\alpha T), & \text{where } \alpha \text{ is a constant,} \\ f_{\Delta\gamma}(T) &= \text{const. exp } (-\beta/T), & \text{where } \beta \text{ is a constant,} \end{aligned}$$

these relations holding above 200° K. All phosphor specimens show a departure from these relations at temperatures below about 200° K, where the dielectric changes are approximately independent of temperature.

No theoretical explanation of these temperature variations of the dielectric changes has been found. Experimental values of the constants α and β in the temperature functions for several phosphors with different luminescence characteristics are given in table 1. An inspection of the table shows that the values of α and β show no marked change from specimen to specimen. It appears from these results that the temperature functions are intrinsic properties of the filled electron traps and do not depend on the particular phosphor to any great extent. Again, the different trap depths associated with long and short phosphorescence specimens have little effect on the constants. Variation of the field frequency appears to have little effect on the temperature functions over a wide range.

TABLE 1. VALUES OF THE CONSTANTS α AND β FOR THE TEMPERATURE VARIATION OF THE DIELECTRIC CHANGES IN PHOSPHORS

phosphor	phosphorescence	α ($^{\circ}\text{K}^{-1}$)	β ($^{\circ}\text{K}$)
ZnS-Zn	very short	0.018	810
ZnS-Ag	short	0.0165	960
ZnS-Cu	short	0.015	1000
ZnS (70 %)-CdS (30 %)-Cu	medium	0.0125	670
ZnS (95 %)-CdS (5 %)-Cu	medium	0.014	720
ZnS-Cu	long	0.010	840

(4) *Comparison of the dielectric changes during fluorescence and thermoluminescence of phosphors*

The ratio of the intensity of fluorescence at room temperature to the maximum intensity of thermoluminescence emission is usually of the order of 100:1 for most specimens. The corresponding ratio for the dielectric changes is of the order of unity. These experimental facts provide further evidence that the dielectric effects are due to trapped electrons and depend on their number, which is not very different during fluorescence or thermoluminescence. The high intensity of fluorescence is due to the return of excited electrons to luminescence centres without being captured by traps between excitation and return. Thus the two ratios would be expected to be of different orders.

Examination of the dielectric changes with temperature during fluorescence shows that their variation is slow. This can be explained as follows: The number of trapped electrons during fluorescence will decrease rapidly with temperature, as

shown by equation (3), for a single trap depth. In addition to this decrease there is the rapid rise in the dielectric effects due to their intrinsic temperature functions described in the last section. The combination of the two different rapid variations with temperature can give rise to a resultant variation with temperature which is relatively slow.

(5) *Variation of the dielectric changes in phosphors with the frequency of the applied field*

Measurements of the behaviour of the dielectric changes in excited phosphors when the frequency of the applied electric field is varied have been made by the use of the circuit magnification meter as stated in an earlier section. It is found that the changes in the dielectric constant due to the filling of electron traps decrease as the frequency of the applied field increases while the changes in dielectric loss increase as the field frequency increases. These variations with field frequency are characteristic of polarizable systems which have finite relaxation times and which therefore show anomalous dispersion and absorption at field frequencies of the same order as the reciprocal of the relaxation times. A general theoretical treatment of such systems has been given by many workers, notably Debye (1929), and in the presentation of the experimental results for phosphors use has been made of the theoretical developments of Murphy & Morgan (1939). Their derivations are for systems in which there is a charged particle of negligible mass bound to some system of relatively high inertia and are found to be applicable to filled electron traps. The present treatment of the results separates the dielectric effects due to filled traps from those occurring in the unexcited phosphor. Control experiments on the unexcited specimens indicate no anomalous variation of the dielectric properties with frequency when all traps are empty. This would be expected as the relaxation times of the matrix ions or atoms of the phosphors are very small, of the order of 10^{-12} sec.

(a) *Variation of the change in dielectric constant with field frequency.* From the derivations of Murphy & Morgan (1939), the dielectric constant change $\Delta\epsilon$ at an angular frequency ω of the applied field is related to its values at static and infinite frequencies $\Delta\epsilon_0$ and $\Delta\epsilon_\infty$ respectively, thus:

$$\Delta\epsilon = \Delta\epsilon_\infty + (\Delta\epsilon_0 - \Delta\epsilon_\infty)/(1 + \omega^2\tau^2),$$

where τ is the relaxation time of the polarizable centre, in this case a filled trap. This equation may be rewritten thus:

$$\frac{\Delta\epsilon_0 - \Delta\epsilon}{\Delta\epsilon - \Delta\epsilon_\infty} = \omega^2\tau^2 = 4\pi^2f^2\tau^2, \quad (8)$$

where f is the field frequency in cyc./sec. In practice $\Delta\epsilon_0$ and $\Delta\epsilon_\infty$ can be obtained by extrapolation at the extreme ends of the frequency range used. Thus the left-hand side of equation (8) should be a linear function of the square of the applied field frequency. Results of experiment for a zinc sulphide-cadmium sulphide phos-

phor activated by copper are shown in figure 11, plotted to show the agreement with equation (8). The ordinate is the expression on the right-hand side of the equation as obtained from experiment, while the abscissa is the square of the field frequency. The results are given for two different temperatures, and it is evident that equation (8) is followed by the experimental results. The slopes of the two straight lines of figure 11 will give values for the relaxation times of the filled traps at the two temperatures. Values obtained in this way for several phosphors are given in column iii of table 2.

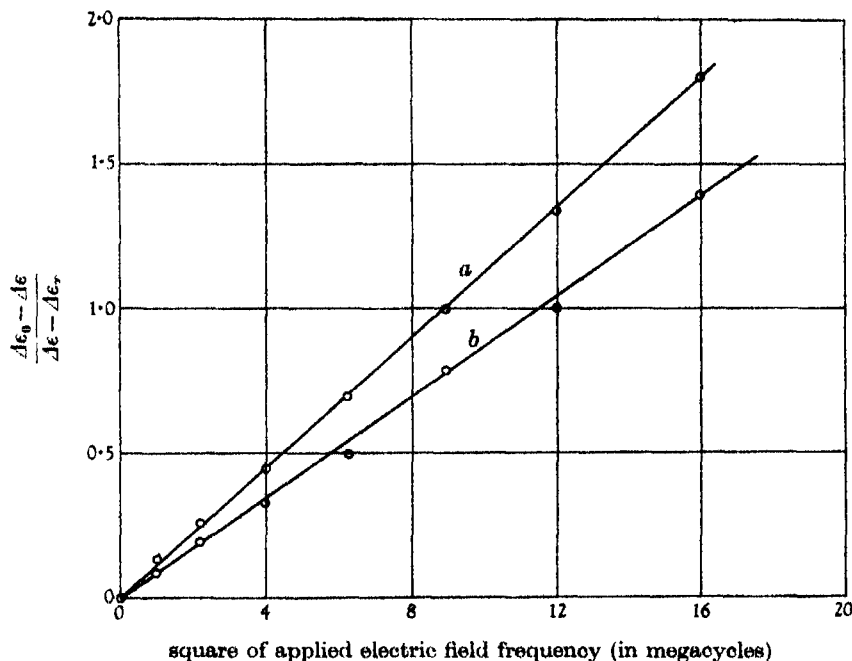


FIGURE 11. Variation of dielectric constant change with applied electric field frequency for ZnS-CdS-Cu phosphor at two different temperatures for constant excitation intensity. (a) Phosphor temperature 291° K; (b) phosphor temperature 361° K. $\Delta\epsilon_0$ = dielectric constant change at zero field frequency. $\Delta\epsilon_\infty$ = dielectric constant change at infinite field frequency. $\Delta\epsilon$ = dielectric change at intermediate frequency.

(b) *Variation of the change in dielectric loss with field frequency.* From the derivations of Murphy & Morgan the change in dielectric loss, expressed as the change in equivalent parallel conductance $\Delta\gamma$, will vary with the applied field frequency according to the relation

$$\Delta\gamma = (\Delta\epsilon_0 - \Delta\epsilon_\infty)\omega^2\tau/3.6\pi \times 10^{12}(1 + \omega^2\tau^2) \text{ ohm}^{-1},$$

which can be rewritten

$$\frac{\tau}{\Delta\gamma} \frac{(\Delta\epsilon_0 - \Delta\epsilon_\infty)}{3.6\pi \times 10^{12}} = \frac{1}{\omega^2} + \tau^2 = \frac{1}{4\pi^2 f^2} + \tau^2. \quad (9)$$

Thus the inverse of $\Delta\gamma$ will vary linearly with the inverse square of the field frequency. Using this equation the experimental results for the same phosphor for which the results are given in figure 11, are presented to show its validity in figure 12. The ordinate is the inverse of the dielectric loss change $\Delta\gamma$ and the abscissa is the inverse square of the applied field frequency. Values of the relaxation times for the filled traps cannot be obtained from these results as the absolute values of $\Delta\epsilon_0$ and $\Delta\epsilon_\infty$ are not known accurately. However, an independent derivation of the values of τ can be obtained from the variation of the loss factor $\Delta\epsilon''$ with field frequency which is given by the relation

$$\Delta\epsilon'' = \frac{4\pi\Delta\gamma}{\omega} = \frac{(\Delta\epsilon_0 - \Delta\epsilon_\infty)\omega\tau}{0.9 \times 10^{12}(1 + \omega^2\tau^2)}.$$

This loss factor will have a maximum when $\omega\tau = 1$. A series of values of τ for different phosphors obtained from the loss factor is given in column iv of table 2.

TABLE 2. EXPERIMENTALLY DERIVED VALUES OF THE RELAXATION TIMES OF FILLED ELECTRON TRAPS

phosphor specimen (i)	tem- perature °K (ii)	relaxation time τ in sec. ($\times 10^{-8}$)	
		(iii)	(iv)
ZnS (70 %)-CdS (30 %)-Cu, orange luminescence, medium phosphorescence	90	9.0	8.0
	291	5.3	7.8
	361	4.7	5.3
	457	4.3	4.0
ZnS-Cu, green luminescence, long phosphorescence	291	13.3	16.0
ZnS-Ag, blue luminescence, short phosphorescence	291	5.8	4.7
CdSe-CdTe-Cu, no luminescence, infra-red excited	90	13.4	15.0
	291	4.4	3.8

(c) *The significance of the relaxation time τ and its variation with temperature.*

Values of the relaxation time τ for filled electron traps in different phosphor specimens have been given in table 2. Included in this table are values for a non-luminescent material which changes its dielectric constant and loss properties when irradiated with visible or infra-red light. The mechanism giving rise to its sensitivity to infra-red is essentially the same as that causing dielectric changes in phosphors, and the values of the relaxation times for this specimen are of the same order as those for phosphors. The values of the relaxation times for different phosphors do not show marked variation from one specimen to another. The results given in the table show that the relaxation time does not vary rapidly with temperature. It is found to be approximately proportional to the inverse of the temperature. The relaxation time of a filled trap is significant, as it should be closely related to the reciprocal of the constant s of equation (1) which is assumed to be a simple function of the natural frequency of vibration of the filled trap (Randall & Wilkins 1945, p. 372). The same workers have made determinations of the value of s for some phos-

phors, but the accuracy of their results is not high enough to make comparisons with similar derivations from the relaxation times of the filled traps determined from the dielectric experiments. In the above derivations of the relaxation times the theoretical relations used neglect the possibility of interaction between adjacent polarized centres. However, such interaction is not likely to affect the value of the relaxation time by a factor greater than two (Murphy & Morgan 1939; Williams 1934; and others). The assumption of non-interaction is probably justified, since the estimates of the concentration of traps in the phosphor matrix give about one trap per 10^3 or 10^4 lattice atoms.

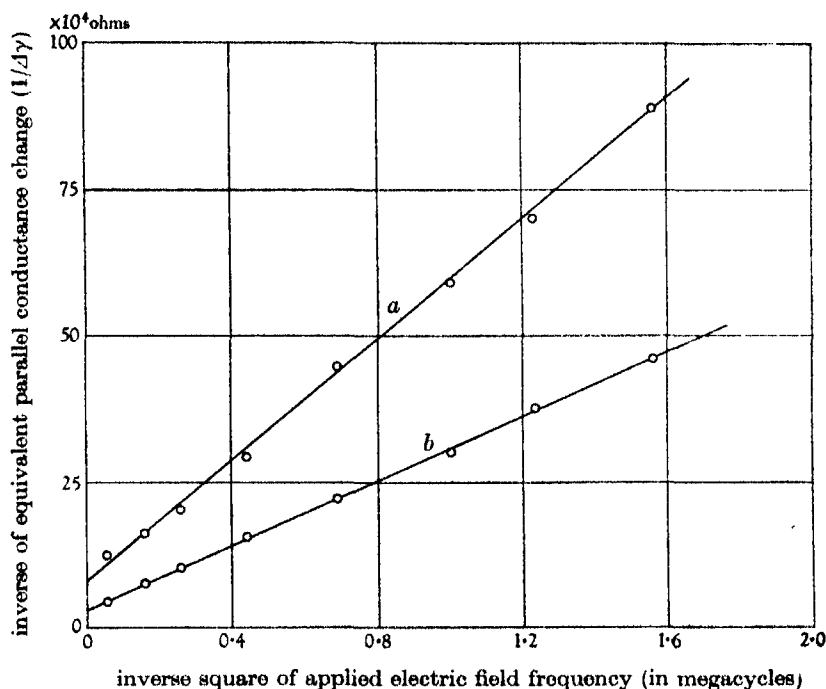


FIGURE 12. Variation of dielectric loss change with applied electric field frequency for ZnS-CdS-Cu phosphor at two different temperatures for constant excitation intensity. (a) Phosphor temperature 291°K; (b) phosphor temperature 361°K.

(6) *Variation of the dielectric changes in phosphors with the wave-length of the exciting radiation*

The effect of the variation of excitation wave-length on the dielectric effects in phosphors has been studied by earlier workers (Gudden & Pohl 1920; Gisolf 1939). The experimental results of these workers have been confirmed. It is found that there are no dielectric changes in phosphors unless the exciting radiation wave-length is sufficiently short to produce photoconductivity, that is, to free electrons from luminescence centres. Thus there are no dielectric effects in zinc sulphide-manganese activated phosphors when excited by radiation of 0.43μ wave-length. This

is explained by the absence of filling of electron traps by such excitation (Garlick & Wilkins 1945). It has been shown by Gisolf (1939) that there is no change in the dielectric loss of such phosphors when the excitation wave-length is in the absorption region of the manganese impurity. In all zinc sulphide phosphors examined the dielectric changes showed the same behaviour when excited by light of wave-lengths between 0.35 and 0.41μ , but they become very small when the wave-length is decreased to less than 0.33μ . This decrease for short wave-lengths can be explained by the rapid rise in absorption coefficient of the phosphors below 0.35μ (Gisolf, de Groot & Kröger 1941). This rise in absorption coefficient allows only a thin layer of the phosphor dielectric to be excited and the saturation of traps occurs at very low intensities of excitation. Thus the dielectric changes per unit volume of the whole dielectric are very small compared with those for excitation with longer wave-lengths, which give almost uniform excitation throughout the phosphor layer. The long wave-length limit of the dielectric phenomenon is dependent upon the constitution of the particular specimen as shown by the special case of the non-luminescent specimen whose relaxation times are given in table 2 for the dielectric effects due to infra-red radiation. In general, it is found that the increase in the percentage of cadmium sulphide in the zinc-cadmium sulphide phosphors shifts the limit for the dielectric phenomenon to longer wave-lengths. This is due to the shift of the wave-length limit for the production of photoconductivity and the filling of electron traps.

CONCLUSION

Studies of the changes in the dielectric constant and dielectric loss of phosphors during excitation by suitable radiation show that the changes are due to the polarization and absorption properties of electron traps which are filled by electrons. The trapped electron can be displaced by small applied electric fields due to its low binding energy compared with the binding energy of electrons bound to lattice atoms or ions or to impurity centres. This displacement constitutes polarization of the filled trap. The changes in dielectric constant and loss observed experimentally depend on the magnitude of the electron displacement and the phase relation of the displacement to the alternating applied field. Phase differences arise because of the finite relaxation times of filled electron traps which are of the order of 10^{-7} sec. Thermoluminescence experiments and measurements of the corresponding dielectric changes show that the polarization and absorption of electron traps containing electrons vary rapidly with the temperature. It is found by experiment that electrons moving in the conduction energy levels of the phosphor crystals make negligible contributions to the dielectric effects compared with those due to filled electron traps. The spectral sensitivity of the dielectric changes has a maximum in the region of 0.35 – 0.41μ . The long wave-length limit of the dielectric phenomenon is dependent upon the particular constitution of each specimen but is the same as that for the production of photoconductivity. These studies provide new and significant additions to the knowledge of the properties of electron traps in phosphors.

Our thanks are due to Professor M. L. Oliphant, F.R.S., for the provision of facilities for this work, to Professor R. Peierls, F.R.S., for much fruitful discussion, and to the Ministry of Supply under whose research projects this work has been carried out. We have received valuable assistance in experimental work from Miss B. M. Browne.

REFERENCES

- Dobye, P. 1929 *Polar molecules*. New York: Chem. Cat. Co.
- Garlick, G. F. J. & Wilkins, M. H. F. 1945 *Proc. Roy. Soc. A*, **184**, 408.
- Gisolf, J. H. 1939 *Physica, Eindhoven*, **6**, 918.
- Gisolf, J. H., Groot, W. de & Kröger, F. A. 1941 *Physica, Eindhoven*, **8**, 805.
- Goos, F. 1939 *Ann. Phys., Lpz.*, **34**, 77.
- Gudden, B. & Pohl, R. 1920 *Z. Phys.* **1**, 365.
- Herweg, J. H. 1923 *Z. Phys.* **16**, 29.
- Lenard, P. & Saeland, S. 1909 *Ann. Phys., Lpz.*, **28**, 240.
- Mott, N. F. & Gurney, R. W. 1940 *Electronic processes in ionic crystals*. Oxford Univ. Press.
- Murphy, E. J. & Morgan, S. O. 1939 *Bell. Syst. Tech. J.* **18**, 502.
- Randall, J. T. & Wilkins, M. H. F. 1945 *Proc. Roy. Soc. A*, **184**, 366.
- Schmidt, F. 1923 *Ann. Phys., Lpz.*, **70**, 16.
- Wilde, R. 1923 *Z. Phys.* **15**, 350.
- Williams, J. W. 1934 *Trans. Faraday Soc.* **30**, 724.

The effect of temperature on the intensity of X-ray reflexion.

By E. A. OWEN AND R. W. WILLIAMS

University College of North Wales, Bangor

(Communicated by Sir Geoffrey Taylor, F.R.S.—Received 3 April 1946)

The effect of temperature on the intensity of X-ray reflexion by gold, copper and aluminium has been studied by making microphotometric measurements on lines in X-ray structure spectra obtained with powder specimens in a Debye-Scherrer camera. A special method was employed to make cylindrical powder specimens, 0.8 mm. in diameter, which held together without adhesive and were free from a core of foreign material. The primary beam was standardized by means of a flat-plate X-ray camera, furnished with a plate of powdered gold which provided a spectrum whose lines could be accurately measured. The powder specimen under investigation was maintained in *vacuo* at temperatures ranging up to about 900° K. and its temperature estimated from lattice parameter measurements.

The observed fall in intensity of X-ray reflexion by gold and copper as the temperature is raised up to about 900° K can be accounted for if the characteristic temperature varies with temperature in accordance with the relation $\Theta_T = \Theta_1[1 - \alpha\gamma(T - T_1)]$, where Θ_T and Θ_1 are the characteristic temperatures at temperatures T and T_1 , α is the cubical coefficient of thermal expansion and γ is the Grüneisen constant. This relation is found to hold also for aluminium up to about 600° K. Beyond 600° K the fall of intensity exceeds that to be expected from the above relation, and it is suggested that another factor becomes prominent in the case of aluminium at the higher temperatures.

The characteristic temperatures of gold, copper and aluminium now found by X-ray measurement at different temperatures, agree with the values obtained at those temperatures by specific heat and electrical conductivity methods.

It was shown in a previous paper (Owen & Edmunds 1945) that, by taking particular care in the preparation of the powder specimen, a fibre X-ray camera may be employed for the investigation of the relative intensity of lines in structure spectra. In that investigation the lines in photographs taken with copper and gold at room temperature were measured. The study of the effect of temperature on the intensity of X-ray reflexion introduces difficulties which were not encountered in measurements made with material at room temperature, the chief of these being the standardization of the X-ray beam, which is necessary before the comparison of reflexions at different temperatures can be made. The measurement of the temperature of the specimen presents no particular difficulty, because with materials the thermal expansions of which have been thoroughly established, this can be accomplished by measuring the lattice parameter. To enhance the accuracy of the measurement of line intensity, special precautions have to be taken in order to obtain structure spectra with as clear a background as possible.

For the photographic basis of the work reference should be made to the paper already referred to; the relation between blackening of the film and the quantity of X-rays falling upon it is there fully investigated. It was found with the technique adopted that it was safe to assume a linear relation between these quantities for values of the blackening up to 0.8 unit; for higher values of blackening the linear relation no longer holds. It was therefore important to keep within the above degree of blackening when making measurements of line intensity.

The microphotometry of lines in the X-ray photographs enables the variation of the intensity with temperature to be determined, and the results can be compared with the prediction of the theoretical formulae developed by Debye & Waller. An important quantity in Debye's theory is the characteristic temperature Θ , and the most instructive way in which the experimental results can be compared with the results of the Debye-Waller theory is to calculate values of the characteristic temperature from them. In this paper we have attempted to do this for gold, copper and aluminium.

EXPERIMENTAL

The apparatus is so designed that the specimen can be maintained, *in vacuo*, at any desired temperature between room temperature and about 600° C. The camera itself remains at room temperature during the exposure. In order to standardize the X-ray beam use is made of the flat-plate camera which is kept at room temperature, and an exposure is made with this camera simultaneously with the fibre camera.

The arrangement of the apparatus is shown diagrammatically in figure 1. The specimen is mounted at *S* and can be rotated continuously during exposure to X-rays by means of a clock mechanism, the housing of which is indicated by the rectangles *P* in the diagram. It is held vertically in a small electrical furnace *F* which is wound in two halves, with a gap between them to allow the middle of the specimen to be exposed to the X-ray beam. The whole is enclosed in the cylindrical

chamber *C*, which can be evacuated. Outside this chamber, the camera *A* is mounted. This can be removed without disturbing the rest of the apparatus, and can be replaced again in exactly the same position on its hole, slot and plane mounting. The X-ray beam was arranged to pass symmetrically through the slit system.

On the other side of the X-ray tube *T* is mounted the flat-plate camera *B*, which receives the X-ray beam passing through the window *W*. The cameras *A* and *B* are situated in the same horizontal plane, and the mounting of the camera *B* is fixed rigidly in relation to the tube, but the camera itself can be removed from its hole, slot and plane mounting and replaced again in the same position without disturbing any other part.

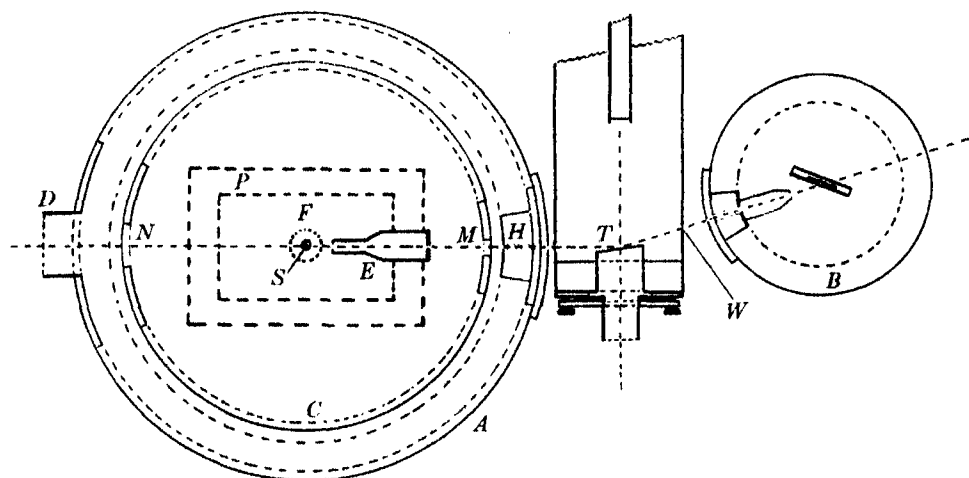


FIGURE 1

The whole assembly to the left of the target (figure 1), including the furnace in its vacuum enclosure and the fibre camera *A*, is mounted on a platform on wheels fitted with roller bearings, and can be moved along rails away from the tube, in order to facilitate manipulation of the camera and its accessories, when a new fibre has to be mounted inside the chamber. It is essential, however, that during a series of exposures at different temperatures with any one specimen, the relative positions of the X-ray tube and the two cameras should be maintained. The construction of the apparatus is such that this condition is satisfied.

To enable the X-rays reflected from the specimen to reach the camera *A*, a slit about half an inch wide, covered with thin cellophane, is made in the cylindrical chamber at the proper height. The entrance and exit holes *M* and *N* in the chamber are also rendered vacuum-tight by means of a cellophane covering. To keep the joint between the cellophane and the brass surface cool, two annular water chambers (not shown in the diagram) arranged within the main chamber, have cold water circulating through them. To absorb water vapour small glass vessels filled with phosphorus pentoxide are placed on the floor of the chamber. The projecting tube

D attached to the camera provides a suitable mounting for a fluorescent screen to find if the specimen is mounted centrally in the X-ray beam. The screen is removed during an exposure to allow a free exit to the beam.

Before commencing an exposure at elevated temperature, a steady current is allowed to flow for half an hour through the heating furnace. That this was a sufficiently long period for the specimen to attain a steady state was proved by the fact that sharp doublets in the precision range of the camera were obtained; also these doublets displayed no dissymmetry such as would result if an appreciable temperature gradient existed in the part of the specimen exposed to the X-ray beam.

The current through the furnace was maintained at a constant value throughout a given exposure. There was no need to calibrate the furnace for temperature because each high-temperature film was made to serve as its own thermometer.

The quality of the films obtained with the fibre camera depends to a great extent on the slit system employed, both as regards the attainment of good resolution and the avoidance of scattered radiation. The slit system is divided into two parts—the hole in the block *H* (figure 1), which is fitted into the camera, and the hole in the block *E* carried by the mounting of the furnace. By taking special measures to shield the film, it was possible to obtain a clear background all along its length. Such background as still existed was mainly due to the occurrence of absorption edges and could, of course, not be eliminated. The mean radius of the fibre camera was 96.21 mm. A series of fiducial marks enabled corrections to be made for film changes on processing.

The preparation of the fibre has been described in the paper already quoted. It requires much patience and care in manipulation to produce a satisfactory fibre, but when made it is accurately cylindrical, holds together without adhesive and is free from any core of foreign material. After the specimen has been annealed to remove the effect of cold work done on the filings, it can be more easily handled and adjusted in the camera mounting; adjustment is necessary so that the specimen may be rendered free from wobble when it is rotated, otherwise the definition of the lines in the photographs is not satisfactory.

Materials of the highest purity were used in the investigation. The size of the crystallites in the specimen is important; only those particles passed by a sieve of 350 mesh and rejected by a sieve made with two layers of 350 mesh gauze, displaced relative to one another through half a diagonal of a mesh, were used. This procedure resulted in some degree of uniformity of particle size. The ideal size of particle is that with which there is no appreciable amount of primary extinction, and at the same time no particle scattering. As in the previous work, particle size was considered to be satisfactory if the lines in photographs obtained when the specimen was not rotated were free from spottiness, and did not display undue broadening. To reduce error arising from background discontinuities, it is also important to obtain well-defined reflexion lines, and this implies the use of a crystalline specimen in good equilibrium, as free from wobble as possible when it is rotated, and a sharp tube focus. All these conditions were striven after and to a great measure secured.

One of the salient features of the present work is the employment of the flat-plate camera to serve as an index of the output of characteristic radiation by the X-ray tube. Need for such an index arises from the fact that to obtain results which can be compared with theoretical prediction, fibre camera exposures at elevated temperatures have to be compared with exposures made at room temperature. No useful information can be obtained from these exposures unless the performance of the tube can be specified for each exposure. In the present investigation the assumption is made that the output of characteristic radiation from the tube divides itself in a constant ratio between the two cameras, and there is ample experimental evidence that this is a legitimate assumption, so long as the target of the tube is not pitted, and precaution was taken that this did not occur.

For details of the flat-plate camera used, reference should be made to the previous paper (Owens & Edmunds 1945). In the preparation of the crystalline powder plate, certain factors have to be attended to in order to obtain lines of uniform intensity. The size of the particles must be small and the distribution of the crystal planes sufficiently random to ensure that the lines are free from spottiness. In the present work a gold flat plate was used throughout; the filings were prepared so as to obtain uniformity of size in the manner already indicated for the fibre specimen, and were annealed before use. The actual surface of gold measured about 10 by 7 mm.

It is to be noted that with the flat-plate camera, a reflected line for which the glancing angle θ is equal to the angle (α) between the beam and the plate, will be sharply focused, but other lines for which θ is not equal to α will be out of focus and of appreciable breadth. At first sight, this characteristic feature of the flat-plate camera appears to be a grave disadvantage, as, indeed, it would be were one concerned with the position of the lines and not with their intensities, but for intensity measurements it is of great value, as in the out-of-focus lines the reflexion is spread out over an appreciable angle and consequently a line can be quite intense without exceeding the limit of the film calibration and can, therefore, be photometered accurately. The accuracy is also improved by the profiles of the out-of-focus lines, which are found to have steep sides and more or less flat tops, and are, therefore, well suited for micro-photometry (Brentano 1935; Brindley & Spiers 1934).

It was necessary to cut down the beam entering the flat-plate camera so that when the longest exposure was made with the fibre camera—which occurred when the specimen was maintained at the highest temperature of which the apparatus was capable—the density of the lines on the flat-plate film would be just within the limit permitted by the calibration curve for the film. The X-ray beam was, therefore, passed through appropriate filters, which also eliminated the β -lines. This procedure makes for maximum accuracy in fixing the intensity level of the exposure.

Ten lines were seen in the photographs obtained with the flat-plate camera, with copper $K\alpha$ radiation, when the value of the angle α is 36° . Of these, three, namely the lines for which $2\theta^2$ is 8, 11 and 12, were selected as the standard lines, the sum of their intensities being taken as proportional to the intensity of the $K\alpha$ radiation passing into the flat-plate camera.

It was arranged that each photograph taken when the fibre was maintained at elevated temperature was bracketed between two photographs taken when the fibre was at room temperature. The intensities of the lines were all brought to the same level by calculating their values per unit intensity of radiation entering the flat-plate camera.

It may be added that in processing the photographic films, precaution was taken that the films from both the fibre and the flat-plate cameras were given identical treatment; they were developed together in a small tank, especially made for the purpose, in which the films were held in position, close together but not touching, thus ensuring that they were developed at the same temperature.

The instrument available for measuring the intensities of the X-ray lines was a Moll, type B, recording microphotometer, which had been modified in the laboratory so that it could be used as a 'null' instrument. This was effected by employing a wedge with a linear calibration curve of density gradation along it. Owing to the sensitivity, stability and excellent dead-beat qualities of the Moll galvanometer, readings with the wedge method could, with confidence, be taken in quick succession. One advantage of the null method over the original recording method is the ease with which check readings of the zero can be made, thus removing uncertainty with regard to galvanometer creep; another advantage is that the curves can be recorded on a magnified scale, which helps to increase the accuracy of measurement. The areas of the peaks were determined by means of a planimeter.

To obtain consistent results in photometering the films it was necessary to pay particular attention to the mounting of the films on the film carriage, so that all the lines were photometered always at the same level. Only lines occurring at large values of glancing angle were measured, for these lines display the greatest change of intensity with change of temperature.

THEORETICAL

The intensity of a line recorded on a powder photograph depends on a large number of factors. In a fibre camera of radius r , suppose there is a volume V of crystalline powder exposed to the incident X-radiation which is of intensity I_0 . Then the diffracted energy P which gives rise to a line of length l on the film is given by the following expression:

$$\frac{P}{I_0} \equiv R = \frac{\lambda^3 e^4}{16\pi m^2 c^4 r} p V N^2 F_T^2 \frac{1 + \cos^2 2\theta}{\sin \theta \sin 2\theta} A = B A V \psi(\theta) N^2 F_T^2. \quad (1)$$

λ is the wave-length of the X-rays; N is the number of unit cells per unit volume of the crystal, and the symbols e , m and c have their usual significance. B is a constant and includes p , the multiplicity factor for the crystal planes (hkl) corresponding to the glancing angle θ , which is a constant since only one line is considered at a time, and $\psi(\theta) = \frac{1 + \cos^2 2\theta}{\sin \theta \sin 2\theta}$. The quantity A is the absorption factor, data for

the calculation of which for a crystalline mass in the form of a cylindrical rod have been tabulated by Bradley (1935).

F_T , the scattering factor at temperature T , refers to the scattering from a unit cell of the crystal, and it is related to the scattering factor at absolute zero, F_0 , by the relation $F_T = F_0 e^{-M}$, where M , in the case of a crystal composed of atoms of one kind arranged at the points of a cubic lattice, is given by the expression

$$M = \frac{6h^2}{mk\Theta} \left\{ \frac{\phi(x)}{x} + \frac{1}{4} \right\} \left(\frac{\sin \theta}{\lambda} \right)^2, \quad (2)$$

where Θ is the characteristic temperature of the crystal and $x = \Theta/T$, T being the absolute temperature. Values of $\phi(x)$ for different values of x have been calculated by Debye. The term $\frac{1}{4}$ inside the bracket is consequent upon the assumption of the existence of zero point energy. In the present work only a very slight difference is introduced into the result if this term is omitted.

Since inquiry is being made into the effect that change in temperature of the crystalline mass will have on the intensity of a line in the powder photograph, the values that various quantities have at the 'standard' room temperature (i.e. 293° K) will be denoted by the suffix 1 and the values at an elevated temperature T by the suffix T . Thus by combining the above relations one arrives at the following working formula:

$$Y \equiv \left(\frac{\lambda}{\sin \theta_T} \right)^2 \left[\log_e \frac{R_T}{R_1} - \log_e \left\{ \frac{A_T V_T \psi(\theta_T) N_T^2 (F_0)_T^2}{A_1 V_1 \psi(\theta_1) N_1^2 (F_0)_1^2} \right\} \right] = - \frac{12h^2}{mk^2\Theta} \left\{ \frac{\phi(x_T)}{x_T} - \frac{\phi(x_1)}{x_1} \right\}. \quad (3)$$

This enables the predictions of theory with regard to the dependence of line intensity on temperature to be compared with the results of experiment. Furthermore, it separates the effect due to increased thermal motion of the atom in the lattice from the effects due to changes in absorption, volume of specimen, glancing angle, and number of unit cells per unit volume, and the slight change in value of F_0 owing to change in angle of reflexion, this being quite apart from the variation of F with temperature as expressed by the relation $F_T = F_0 e^{-M}$. These changes may be regarded as corrections which must be applied to the experimentally determined ordinate before this can be compared with the results of theory. Since the term on the right-hand side of the above relation does not involve θ , it is clear that if at a given temperature the corrected expression $(\lambda/\sin \theta_T)^2 \log R_T/R_1$ is evaluated for various lines on the film, these values should be constant. This was found to be the case within experimental error.

In the above relations the characteristic temperature is assumed to be independent of the temperature. If it varies with temperature one may proceed as follows. The crystal lattice alters its dimensions with change of temperature, and if the Einstein model of a solid is assumed, an expression can be obtained for the thermal coefficient of expansion, which may be thrown into the following form (Mott & Jones 1936):

$$\frac{\alpha V_0}{\chi_0 C_v} = - \frac{d(\log \nu)}{d(\log V)} \equiv \gamma,$$

when α is the coefficient of thermal expansion, V_0 the volume of 1 g. of the solid, χ_0 the compressibility at zero temperature, C_v the specific heat, and ν the frequency of the atoms in the lattice. γ is the Grüneisen constant and is independent of temperature; its value for a given material may be calculated from the experimental values of α , χ_0 , etc., at any convenient temperature. For gold, copper and aluminium, it has the values 3.03, 1.96 and 2.17 respectively.

Since, in the Einstein distribution function, $h\nu = k\Theta$, it follows that

$$d(\log \Theta) = -\gamma d(\log V),$$

so that

$$\Theta = \text{const. } V^{-\gamma}.$$

Since

$$V_T = V_1[1 + \alpha(T - T_1)],$$

then

$$\Theta_T = \Theta_1[1 - \alpha\gamma(T - T_1)], \quad (4)$$

where Θ_T and Θ_1 are the characteristic temperatures at T and T_1 .

Since Θ now varies with temperature, then

$$Y_1 = -\frac{12h^2}{mk} \left\{ \frac{1}{\Theta_T} \frac{\phi(x_T)}{x_T} - \frac{1}{\Theta_1} \frac{\phi(x_1)}{x_1} \right\}.$$

Using the approximate form $(1 - K_1x)$ for $\phi(x)$, where K_1 is a constant, substituting the value of Θ_T in terms of Θ_1 , and neglecting terms of second order, then, finally,

$$Y_1 = -\frac{12h^2}{mk} \frac{1}{\Theta_1^2} (T - T_1) (1 + 2\alpha\gamma T).$$

When the values of Y_1 given by this relation are plotted against the absolute temperature, we get a curve which is concave to the temperature axis.

If the characteristic temperature were independent of temperature, then, on the above assumptions,

$$Y = -\frac{12h^2}{mk} \frac{1}{\Theta_1^2} (T - T_1),$$

so that

$$Y_1 = Y(1 + 2\alpha\gamma T). \quad (5)$$

There is a small correction to be made owing to the variation of α with temperature. This relation can be applied to the experimental results now obtained either by calculating values of Y from the observed values of the experimental ordinate at different temperatures, and finding if they fall on a straight line, or otherwise by assuming a definite constant value of the characteristic temperature, calculating values of Y_1 at different temperatures and finding how well they fit the experimental curve.

RESULTS

The mean values of the corrected experimental ordinate at different temperatures for gold, copper and aluminium obtained in the present investigation are given in table 1. The probable error of these values is estimated to be about $\pm 5\%$.

TABLE 1

gold		copper		aluminium	
temp. °K	corrected experimental ordinate ($Y_1/2.303$)	temp. °K	corrected experimental ordinate ($Y_1/2.303$)	temp. °K	corrected experimental ordinate ($Y_1/2.303$)
293	0.0	293	0.0	293	0.0
453	-0.26 ₆	398	-0.10 ₆	419	-0.34 ₇
591	-0.62 ₃	453	-0.33 ₈	541	-0.73 ₆
673	-0.66 ₄	517	-0.42 ₀	613	-0.82 ₅
755	-0.83 ₃	585	-0.53 ₈	685	-1.03 ₁
838	-1.21 ₂	652	-0.67 ₆	757	-1.54 ₄
921	-1.22 ₁	717	-0.72 ₁	829	-1.82 ₉
		785	-0.97 ₀	903	-2.38 ₃
		851	-0.94 ₆		

Gold

With gold there was little risk of surface contamination at high temperatures. Four lines only, corresponding to the highest glancing angles, were photometered for intensity measurements, but each side of the photograph was treated separately, thus yielding eight intensity measurements for each exposure. Knowing the temperature, the various corrections were readily calculated and the corrected experimental ordinate for each reflexion obtained. The values for each temperature were reasonably constant as expected by theory. In figure 2 the curve *A* is the straight line obtained on the assumption that the characteristic temperature of

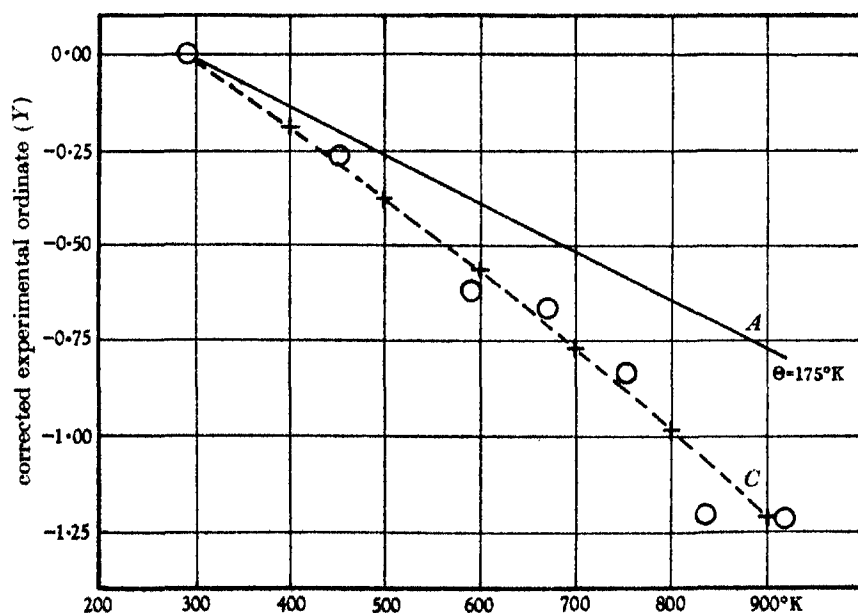


FIGURE 2. Gold.

gold is 175°K and that it is independent of temperature. The open circles represent the observed points, and the crosses are the points calculated from curve *A* by using relation (5). The dotted curve passes through the calculated points, and it is also the best curve that can be made to fit the experimental points. It may therefore be concluded that the characteristic temperature of gold at room temperature (i.e. 293°K) is 175°K , and that its value at other temperatures, up to 900°K , may be calculated by means of relation (4).

Copper

It was important when working with copper that the highest vacuum was maintained and that there was freedom from moisture. The conditions of experiment were such that after many exposures at high temperatures the copper showed but slight tarnish. Several specimens were employed in the course of the investigation, and the results obtained with these were consistent one with the other.

The experimental values are represented by the open circles in figure 3. In the preliminary stages a smooth curve was drawn through these points, and from this curve the values of the experimental ordinate, assuming the characteristic temperature to be independent of temperature, were calculated by means of relation (5). A rough idea was thus obtained of the value of the characteristic temperature at room temperature. The best value was found to be 314°K . Calculating Y_1 from the straight line corresponding to $\Theta = 314^{\circ}\text{K}$, the points marked with crosses were obtained, and the curve through these points is again the best curve through the experimental points. Thus the experimental determination of the ordinate supports the conclusion that the characteristic temperature obeys relation (4), and that the value of the characteristic temperature for copper at 293°K is 314°K .

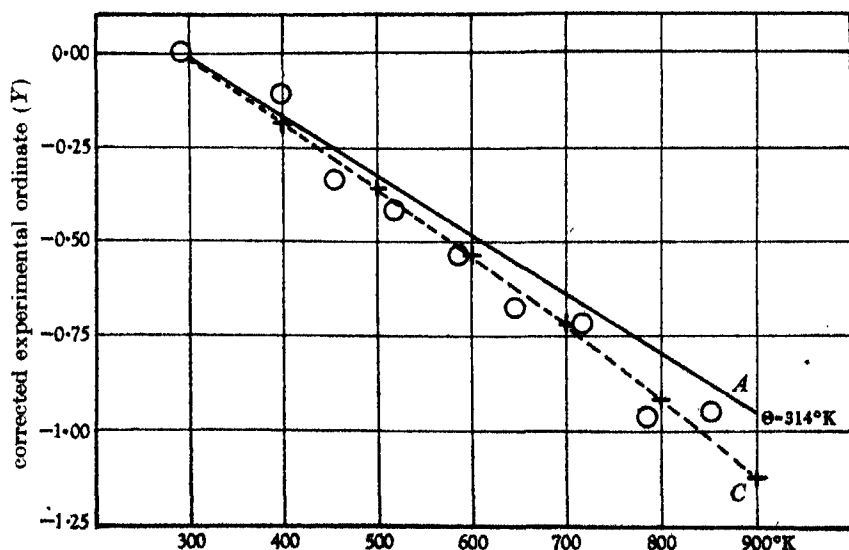


FIGURE 3. Copper.

Aluminium

Measurements on the variation of the intensity of X-ray reflexion by aluminium with change of temperature have already been carried out over the range from liquid-air to room temperature (James, Brindley & Wood 1929). Also several specific-heat determinations have been made, especially at low temperatures.

A single series of exposures was made comprising eight room-temperature exposures and seven at elevated temperatures, and the experimental conditions were favourable for accurate determinations of the experimental ordinate. The copper target used in the X-ray tube for the other two elements examined, was, with aluminium, changed for a cobalt target, in order to secure lines favourably placed for accurate parameter determinations. To standardize the X-ray beam lines in the flat-plate camera, photographs for which the values of Σh^2 were 4, 8 and 11 proved suitable.* In exposures at the highest temperatures line 4 was almost invariably too heavily exposed, thus invalidating the calibration law of the film. In these cases lines 8 and 11 were found to be quite adequate to fix the intensity of the primary beam.

The values of the experimental ordinate corrected for all errors are included in table 1. When plotted they fall on a curve (curve *B* in figure 4) which exhibits a more pronounced curvature than the curves obtained with gold and copper. The value of the experimental ordinate shows a rapid fall in the region of the highest temperatures reached.

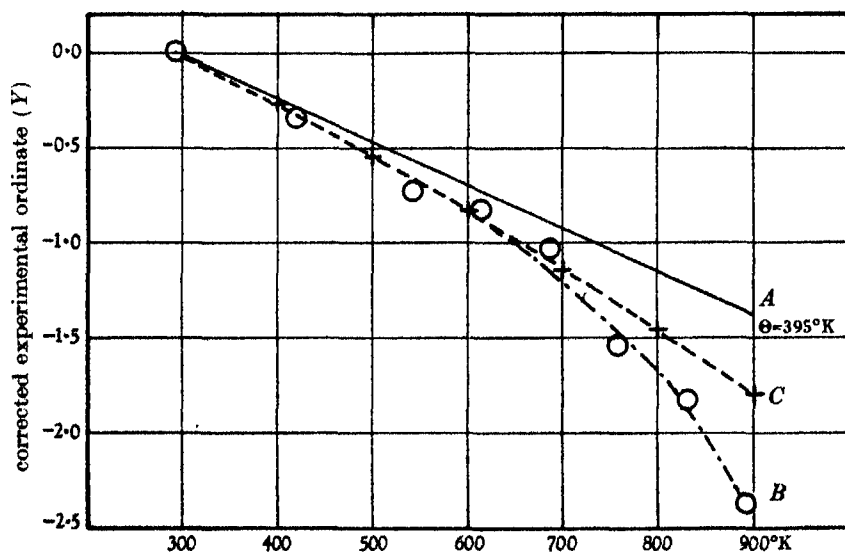


FIGURE 4. Aluminium.

Proceeding along the same lines as before, a value of 395° K for the characteristic temperature of aluminium at room temperature yields the curve *C* when relation (5) is employed. The curve follows curve *B* from room temperature up to about

600° K, but beyond this temperature the values of the experimental ordinate decrease much more rapidly than the calculated values do. From the agreement between the observed and the calculated values of the experimental ordinate from 293 and 600° K, on the assumption that at room temperature the characteristic temperature of aluminium is 395° K, it may be concluded that over this restricted range the characteristic temperature of aluminium obeys relation (4) as do gold and copper.

The rapid fall in intensity of the reflected beam as the temperature of aluminium is raised beyond 600° K requires explanation. It is possible that this decrease of intensity may be due to the removal of energy from the primary beam to form extra reflexions, the intensities of which become very strong in the case of aluminium at the higher temperatures.

The foregoing results show that the variation of characteristic temperature with temperature is represented satisfactorily by relation (4) over the range of temperature investigated. This relation cannot hold for temperatures below a certain limit, which will vary for different elements, but it was of interest to find how far the range over which it would be applicable could be extended below room temperature. Accordingly, it was employed to calculate the characteristic temperatures at the mean temperatures used in some previous investigations at low temperatures. These values are included in table 2 together with the experimental values, and, in the few instances cited, it is found that there is fair agreement between them. It would appear that the relation holds for gold and copper down to about 100° K and for aluminium down to about 10° K.

TABLE 2

metal	mean temp. ° K	characteristic temperature (° K)			
		calculated	observed	method	authority
Au	94	180	186	s.h.	Clusius & Harteck (1928)
Au	175	178	175	e.c.	see Barden (1940)
Cu	100	320	320	s.h.	see Barden (1940)
Al	10	414	419	s.h.	Kok & Keesom (1937)
Al	190	402	403	X-ray	James, Brindley & Wood (1929)

s.h. = specific heat. e.c. = electrical conductivity.

TABLE 3

temp. ° K	characteristic temperature (° K)		
	gold	copper	aluminium
10	—	—	414
100	180	320	408
293	175	314	395
400	173	311	388
600	168	304	374
900	162	295	—

The results of this paper may be summarized by giving the values of the characteristic temperatures of the three elements gold, copper and aluminium, at different temperatures, basing these on the room-temperature values found in the course of the investigation and employing relation (4) to find them at other temperatures. They are collected in table 3. The decrease in the value of the characteristic temperature as the temperature is raised, is 2.2, 3.1 and 6.8° K per 100° K rise of temperature for gold, copper and aluminium respectively.

We take this opportunity of thanking the Department of Scientific and Industrial Research for a grant which enabled the work to be carried out.

REFERENCES

- Barden, J. 1940 *J. Appl. Phys.* **11**, 88.
Bradley, A. J. 1935 *Proc. Phys. Soc.* **67**, 879.
Brentano, J. C. M. 1935 *Proc. Phys. Soc.* **47**, 932.
Brindley, G. W. & Spiers, F. W. 1934 *Proc. Phys. Soc.* **46**, 841.
Clusius, K. & Harteck, P. 1928 *Z. phys. Chem. A*, **134**, 243.
James, R. W., Brindley, G. W. & Wood, R. G. 1929 *Proc. Roy. Soc. A*, **125**, 401.
Kok, J. A. & Keesom, W. H. 1937 *Physica, Eindhoven* **4**, 835.
Mott, N. F. & Jones, H. 1936 *Properties of metals and alloys*, p. 19. Oxford: Clarendon Press.
Owen, E. A. & Edmunds, I. G. 1945 *Phil. Mag.* **36**, 54.

On the theory of dielectric breakdown in solids†

By H. FRÖHLICH, *H. H. Wills Physical Laboratory, University of Bristol*

(Communicated by N. F. Mott, F.R.S.—Received 6 April 1946)

It is shown that the theory of dielectric breakdown in solids previously developed by the author is correct only below a critical temperature T_c . This temperature is defined in such a way that above T_c the density of electrons (in strong fields) is so high that mutual collisions between electrons are more frequent than collisions between electrons and the lattice vibrations. In the presence of strong external fields this leads to an equilibrium distribution of the electrons at an electronic temperature T which is higher than the lattice temperature T_0 . T is determined by the energy balance according to which the rate of energy transfer from the field to the electrons must be equal to the rate of energy transfer from the electrons to the lattice vibrations. It is shown that equilibrium can be obtained only if the field is below a critical field F^* . For stronger fields the electronic temperature T rises steadily until the crystal breaks down. It is found that F^* decreases exponentially with increasing lattice temperature.

The theory now accounts for the rise of dielectric strength with temperature at low temperatures (previous theory) and for its decrease at high temperatures. It also shows why influences which tend to increase the dielectric strength at low temperatures (e.g. admixture of foreign atoms) tend to decrease it in the high-temperature region.

The increase of the electronic temperature with the field strength F leads (for $F < F^*$) to an increase of electronic conductivity with F which is calculated quantitatively.

† Based on Report L/T 153 of the British Electrical and Allied Industries Research Association (E.R.A.).

INTRODUCTION

The theory of dielectric breakdown in ionic crystals developed in recent years by the author (Fröhlich 1937, 1939, 1941, 1942; also Whitehead 1945) succeeded in describing the main features of this phenomenon up to a temperature T_c where a change in the type of breakdown occurs. As one of the main results this theory predicted the positive temperature gradient of the breakdown strength which subsequently was found experimentally by Austen, Hackett & Whitehead (1939, 1940) and, independently, by Buehl, Maurer and von Hippel (cf. Buehl & von Hippel 1939; von Hippel & Maurer 1941). Above T_c , however, the temperature gradient changes sign. Originally this was taken as an indication that breakdown through thermal instability had started, but experimentalists maintained that this was not the case (von Hippel & Lee 1941). Moreover, careful experiments by Thomas & Griffith (1942) on amorphous substances yielded this negative temperature gradient down to temperatures which are sufficiently low to rule out thermal breakdown. It is thus desirable to investigate theoretically why the theory should become invalid in the high-temperature range beyond T_c and to develop a theory for this range. This will be carried out in the present paper. Amorphous substances will be given special consideration in view of their low critical temperature T_c .

To find the reason for the failure of the former theory at high temperatures it should be remembered that breakdown was assumed to be due to an instability of the electrons in the conduction levels of an insulator. From elementary considerations a condition for the occurrence of this instability had been derived which yielded the breakdown strength in terms of the mean free path of those electrons whose energy I is just sufficient to carry out an ionization of the ions or atoms of the lattice. An exact proof for this condition has not been given so far, but the accompanying paper (Fröhlich 1947) gives strong support for its approximative validity by showing that the possibility of reaching stationary conditions is determined by the behaviour of electrons with energy I . To derive this it was required to assume that mutual collisions between electrons can be neglected. This assumption can be justified at sufficiently low temperatures where the density of electrons in conduction levels is very small, although it should be recognized that the density in strong fields—just below breakdown—which is required here may be much higher than the density in weak fields. In any case, however, the density will increase with temperature and at a certain temperature will thus reach a value where the above assumption becomes invalid. It will be seen that this marks the change over in the type of breakdown.

The different behaviour of electrons in this high-temperature region is best demonstrated on the energy exchange between the electrons on the one side, and the external field and the lattice vibrations on the other. At low temperatures, because of the small electronic density, each electron can be treated on its own. At sufficiently high temperatures, however, the electrons, in view of their mutual collisions, will exchange energy with each other more rapidly than with either the field or the lattice vibrations. Thus an external field may be considered to raise the temperature of the

electrons to a value T which is higher than the lattice temperature T_0 , and which depends on the field strength. It has then to be investigated if equilibrium can be reached in this way. This will be found to be possible below a critical field strength.

The increase of electronic temperature with field strength clearly leads to deviations from Ohm's law in strong fields. Thus the following investigations will lead to a theory of the deviations from Ohm's law as well as to a theory of breakdown.

It will be found that the electronic temperature T , though higher, has always the same order of magnitude as the lattice temperature T_0 . In the high-temperature region, therefore, the breakdown field is determined by the properties of slow (thermal) electrons in contrast to low-temperature breakdown which depends on the properties of fast electrons whose energy is of the order I , i.e. several eV. Obvious experimental evidence for this can be found in the increase of breakdown strength in the low-temperature region with temperature and with foreign admixtures. Qualitatively, this resembles the behaviour of the electrical resistance of metals, and both cases are explained by the properties of the mean free path of fast electrons which decreases with increasing temperature, or with increasing concentration of foreign atoms. Slow electrons cannot be expected to behave in this way and, in fact, the breakdown field in the high-temperature region behaves in the opposite way. There have been discussions in the past as to whether it is an instability of the fast or of the slow electrons which is responsible for breakdown. The present investigation leads to a compromise on this question: slow electrons by co-operative action are responsible for breakdown in the high-temperature region, whereas independent fast electrons are responsible for breakdown in the low temperature region.

DEVELOPMENT OF THE THEORY

Consider a solid insulator or semi-conductor in the presence of an external electric field F . A number of electrons, say N_1 , will be in the conduction levels of the solid and give rise to an electric current whose density j is given by

$$j = \sigma F = \frac{e^2 \tau_1 N_1}{m} F, \quad (1)$$

where σ is the conductivity, and τ_1 is the average time of relaxation. This current leads to an energy transfer from the field to the electrons at a rate of

$$A = jF \quad (2)$$

per unit volume. To obtain a stationary state the electrons must be able to transfer energy to the lattice vibrations at the same rate.

The main assumption to be made for the following calculations is to consider the probability for collisions amongst electrons to be large compared with the probability for collisions between electrons and lattice vibrations. The electrons can thus be assumed to be in thermal equilibrium corresponding to an electronic temperature T . This temperature must be higher than the lattice temperature T_0 because otherwise

no energy would be transferred to the lattice. The temperature T has to be determined from the energy balance

$$A = B(T, T_0), \quad (3)$$

where $B(T, T_0)$ is the rate of energy transfer per unit volume from the electrons to the lattice vibrations.

In order to calculate this energy transfer it will be of importance to consider not only electrons in conduction levels but also electrons which are trapped in energy levels due to lattice imperfections. Such levels will be denoted as isolated levels because they lead to a binding of electrons in the neighbourhood of a lattice imperfection. In pure ionic crystals lattice imperfections are due to ions in interstitial positions (cf. Mott & Gurney 1940), but in amorphous substances further imperfections of a different nature should be expected. Isolated levels consist in general of a ground-level and of a number of excited levels (cf. Mott & Gurney 1940) leading up to the conduction levels of the lattice. The distance between the ground-level and the first excited level is usually of the order of 1 eV and will be denoted by $2(V - \Delta V)$. The energy range of the excited levels below the conduction levels will be denoted by ΔV as indicated in figure 1. The average number $f(E/kT)$ of electrons with an energy E above the first excited level is thus (cf. books on semi-conductors) given by

$$f(E/kT) \propto e^{-(V - \Delta V)/kT - E/kT}. \quad (4)$$

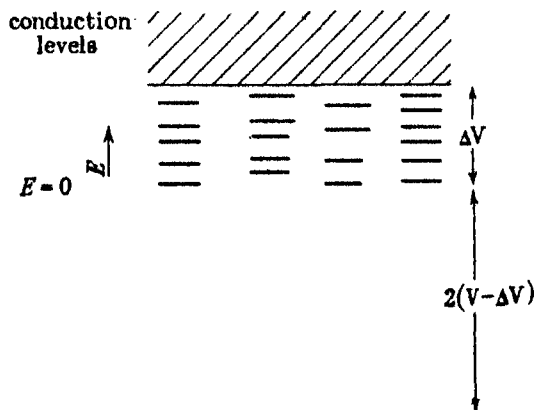


FIGURE 1. Energy levels.

Hence the total number N_1 of electrons in the conduction levels ($E > \Delta V$) is found to be

$$N_1 = \sum_{E > \Delta V} f(E/kT) = C_1(T) e^{-V/kT}, \quad (5)$$

where $C_1(T)$ is a quantity which compared with the exponential term varies slowly with T . Furthermore, the number of electrons in excited isolated levels ($0 < E < \Delta V$) is approximately given by

$$N_2 = \sum_{E=0}^{\Delta V} f(E/kT) \simeq C_2(T) e^{-V/kT + \Delta V/kT} \quad (6)$$

if one assumes that

$$\Delta V \gg kT. \quad (7)$$

Again $C_2(T)$ varies slowly with temperature. The ratio C_1/C_2 is roughly equal to the ratio of the number of energy levels in the interval kT of the conduction levels to the corresponding number in the isolated levels. Denoting the latter per unit volume by z , and considering the electrons in the conduction levels as free electrons one finds approximately

$$\frac{C_1}{C_2} \approx \frac{1}{z} \left(\frac{mkT}{h^2} \right)^{3/2},$$

i.e. for $T \approx 300^\circ$ abs., $C_1/C_2 \approx 10^{19}/z$. This, and (7), allows one to conclude that

$$N_2 \gg N_1, \quad (8)$$

which, according to (5) and (6) and the above estimate, requires that

$$e^{\Delta V/kT} \gg C_1/C_2 \approx 10^{19}/z \quad \text{or} \quad z \gg 10^{19} e^{-\Delta V/kT}.$$

This it seems will be fulfilled not only for amorphous substances but also for crystals unless special care for their purity and perfection has been taken. For assuming, for example, $\Delta V/kT = 5$ (consider (7)), one requires $\Delta Vz/kT > 10^{18}$, i.e. in a solid with 10^{22} atoms per c.c. there should be more than one centre for isolated levels per 10^4 atoms.

It will now be assumed that electrons in the conduction levels as well as in isolated levels can exchange energy with the lattice vibrations. For the sake of simplicity all lattice vibrations are considered to have the same frequency ν . Electrons with energy E may thus either absorb or (provided $E > h\nu$) emit a quantum $h\nu$ and make a transition into the level $E \pm h\nu$. This requires, of course, that the excited isolated levels are sufficiently dense, and that

$$\Delta V \gg h\nu. \quad (9)$$

Transitions from excited isolated levels to their ground-level can be neglected for they would require the simultaneous emission of a great number of quanta $h\nu$.

The above considerations are sufficient to allow a calculation of the temperature dependence of the energy transfer A and B . A quantitative calculation cannot, however, be attempted at present, for our knowledge of the quantitative behaviour of slow (thermal) electrons is not yet sufficiently developed.

Since electrons in isolated levels cannot (or can only very slowly) move through the lattice, it follows that only the N_1 electrons in the conduction levels contribute to the energy transfer A from the field to the lattice. Thus from (2), (1) and (5) one finds

$$A = \frac{e^2 \tau_1 C_1}{m} F^2 e^{-V/kT}. \quad (10)$$

On the other hand, in view of (8), the energy transfer B from the electrons to the lattice is mainly due to the N_2 electrons in isolated levels. The number of transitions per second of electrons from the energy E into $E + h\nu$, connected with absorption of a quantum $h\nu$ is given by

$$W_1 = P(E) f(E/kT) n(T_0), \quad (11)$$

where $P(E)$ is a transition probability which is a function of E but is independent of temperature, and $n(T_0)$ is the number of quanta present at the lattice temperature T_0 , i.e.

$$n(T_0) = \frac{1}{e^{h\nu/kT_0} - 1} \quad \text{and} \quad 1 + n(T_0) = e^{h\nu/kT_0} n(T_0). \quad (12)$$

Similarly, using (4) and (12), the number of inverse processes is given by

$$\begin{aligned} W_2 &= P(E) f((E + h\nu)/kT) (1 + n(T_0)) \\ &= P(E) f(E/kT) n(T_0) e^{h\nu/kT_0 - h\nu/kT}. \end{aligned} \quad (13)$$

In thermal equilibrium in the absence of a field $T = T_0$, and one finds $W_1 = W_2$ as required. From (11) and (13) the total rate of energy transfer is found to be

$$\begin{aligned} B(T, T_0) &= h\nu \sum_{E=0}^{\Delta V} (W_1 - W_2) \\ &= h\nu n(T_0) [e^{h\nu/kT_0 - h\nu/kT} - 1] \sum P(E) f(E/kT). \end{aligned}$$

Let $1/\tau_s$ be the average value of $P(E)$, i.e. (cf. (4) and (6))

$$\frac{N_s}{\tau_s} = \sum P(E) f(E/kT).$$

Again considering (4) and (6) it may be assumed that τ_s varies only slowly with T . Thus inserting N_s from (6) one finds

$$B(T, T_0) = \frac{h\nu}{\tau_s} C_s n(T_0) e^{-V/kT + \Delta V/kT} [e^{h\nu/kT_0 - h\nu/kT} - 1]. \quad (14)$$

Then introducing (14) and (10) into the energy balance (3) yields after division by $h\nu n(T_0) N_s/\tau_s$

$$DF^2 e^{-\Delta V/kT} = e^{h\nu/kT_0 - h\nu/kT} - 1, \quad (15)$$

where

$$D = \frac{e^2 \tau_1 \tau_s C_1}{m h \nu C_s n(T_0)} \quad (16)$$

can be considered as varying slowly with T and will be treated as a constant. Equation (15) will be used to find the electronic temperature T in terms of the field F which will be treated as a parameter. Figure 2 shows the T -dependence of both sides of (15). Apart from practically constant factors the left-hand side represents the rate of energy transfer from the field to the N_1 conduction electrons after division by the number N_s of electrons in the excited isolated levels. It is thus proportional to F^2 , and its temperature dependence is determined by the ratio N_1/N_s of the number of electrons in the conduction levels to the number of electrons in excited isolated levels, which increases with T . The right-hand side of (15), again apart from unimportant factors, represents the rate of energy transfer from the electrons to the lattice divided by the total number of electrons. It vanishes at $T = T_0$, then rises steadily and approaches a constant value as $T \rightarrow \infty$.

Figure 2 shows that for small values of the field F , equation (15) has two solutions for the electronic temperature T ; the first $T = T_1$ is near T_0 and represents a stable equilibrium, while the second one $T = T_2$ lies at very large values of T and leads to an unstable state. On application of a field, energy is supplied to the conduction electrons which make frequent mutual collisions and collisions with the trapped electrons. Thus the electronic temperature rises until it reaches the equilibrium value T_1 of T , i.e. the lower solution of (15). Now if the field F is increased T_1 increases, and T_2 decreases. A critical field $F = F^*$ exists where the two solutions meet, i.e. where $T_1 = T_2 = T^*$. For $F > F^*$ no solution exists because the rate of energy transfer A from the field to the electrons is always larger than the rate of energy transfer B from the electrons to the lattice. Thus the electron temperature T will grow indefinitely until the lattice breaks down. Hence F^* is the breakdown field.

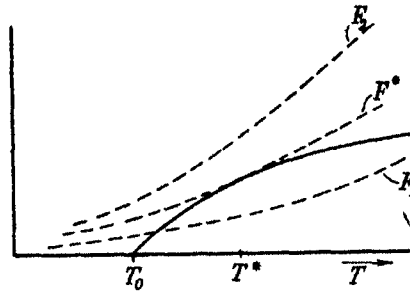


FIGURE 2. Solution of equation (15). — Right-hand side of (15), proportional to energy loss per electron to the lattice. ---- Left-hand side of (15) proportional to energy gain per electron from the field F for three different values F_1 , F^* , F_2 of F . F^* is the breakdown field.

To obtain F^* equation (15) must be fulfilled for $F = F^*$, and in addition the two solutions T_1 and T_2 must be equal. It can be seen from figure 2 that this is fulfilled if the derivatives of both sides of (15) with respect to T are equal, i.e. if

$$DF^* e^{-\Delta V/kT} \Delta V = e^{\hbar\nu/kT_0 - \hbar\nu/kT^*} \hbar\nu. \quad (17)$$

The solution $T = T^*$ of (17) and of (15) (with $F = F^*$) must thus satisfy (eliminate F^* from (17) and insert into (15))

$$e^{\hbar\nu/kT_0 - \hbar\nu/kT^*} \left(1 - \frac{\hbar\nu}{\Delta V}\right) = 1,$$

or in view of (9)

$$\frac{\hbar\nu}{kT_0} - \frac{\hbar\nu}{kT^*} \simeq \frac{\hbar\nu}{\Delta V}. \quad (18)$$

Inserting this into (17) the breakdown field is found to be

$$F^* = ce^{\Delta V/nkT_0}, \quad (19)$$

where using (9)

$$c^2 \simeq \frac{1}{eD} \frac{\hbar\nu}{\Delta V} \quad (e = 2.718...) \quad (20)$$

can (cf. (7) and (9)) be considered to vary only slowly with T_0 . The electronic temperature T is only slightly higher than the lattice temperature T_0 as can be seen from (18) which with the use of (7) yields

$$\frac{T^* - T_0}{T^*} = \frac{kT_0}{\Delta V} \ll 1.$$

The number $N_1(T^*)$ of electrons in the conduction levels may, however, be much larger than without a field ($N_1(T_0)$) for from (5) and (18) it follows that

$$\frac{N_1(T^*)}{N_1(T_0)} \simeq e^{-V/kT^* + V/kT_0} = e^{V/\Delta V}, \quad (21)$$

which may be very large.

Consider finally the case $F < F^*$ in which a stationary state with electronic temperature T_1 will be reached. T_1 depends on the field F which means that the conductivity σ depends on the field strength as can be seen from (1) and (5) according to which $\sigma \propto e^{-V/kT}$. Thus denoting by σ_0 the conductivity for very weak fields where $T_1 \rightarrow T_0$,

$$\sigma = \sigma_0 e^{-V/kT_1 + V/kT_0}. \quad (22)$$

In very strong fields, just below breakdown, $T_1 \simeq T^*$, and hence with (18) or (21),

$$\log \frac{\sigma^*}{\sigma_0} = \frac{V}{\Delta V}, \quad F \lesssim F^*. \quad (23)$$

On the other hand if $F^2 \ll F^{*2}$ the right-hand side of (15) may be developed into

$$DF^2 e^{-\Delta V/kT_1} = \frac{h\nu}{kT_0} - \frac{h\nu}{kT_1}.$$

Using (19) and (20) one obtains

$$\frac{F^2}{F^{*2}} e^{\Delta V/kT_0 - \Delta V/kT_1} \simeq \frac{F^2}{F^{*2}} \simeq \left(\frac{\Delta V}{kT_0} - \frac{\Delta V}{kT_1} \right) e,$$

or with (22)
$$\log \frac{\sigma}{\sigma_0} = \frac{F^2}{F^{*2}} \frac{V}{\Delta V} \frac{1}{e}, \quad F^2 \ll F^{*2} \quad (e = 2.718...). \quad (24)$$

DISCUSSION

The preceding considerations lead to the conclusion that two different types of intrinsic dielectric breakdown have to be distinguished. They are characterized by the temperature dependence of the breakdown field F^* . In the low temperature region the theory developed in previous papers (Fröhlich 1937, 1942) holds, which means that F^* increases with increasing lattice temperature T_0 (or it may be practically constant if the frequencies of the lattice vibrations are sufficiently high). In the high-temperature region, on the other hand, the theory developed in the previous section

of the present paper holds, which means that F^* decreases with increasing temperature according to the law (19) which may also be written as

$$\log F^* = \text{constant} + \frac{\Delta V}{kT_0} \quad (T_0 > T_c). \quad (25)$$

T_c is a temperature which allows to distinguish between the two cases although the transition will not be an abrupt one. The temperature T_c is defined in such a way that in strong fields just below breakdown the density of electrons in the conduction levels is just sufficient to make mutual collisions between electrons as frequent as collisions of electrons with the lattice vibrations. Below T_c collisions between electrons can be neglected whereas above T_c they are predominant. Thus in the low temperature region the behaviour of single electrons has to be considered. Equation (20) of the accompanying paper (Fröhlich 1947) shows that in strong fields this leads to the existence of a considerable number of fast electrons. In this case the breakdown field is determined by the mean free path of fast electrons. This can be calculated quantitatively and a quantitative calculation of the breakdown field is, therefore, possible.

In the high-temperature region the collective behaviour of the electrons has to be considered. This is largely determined by the properties of slow thermal electrons. In this case it has not been possible, so far, to obtain quantitative theoretical values for the breakdown field. The main check of this theory, therefore, lies in a comparison of the temperature dependence (25) with experiments.

Before this is carried out some striking qualitative features of the theory should be mentioned. No attempt has been made to calculate the transition temperature T_c , but it is evident that T_c should be the lower the larger the number of centres for trapped electrons, because T_c is the temperature at which the electron density in the conduction levels attains a certain critical value. Thus admixture of foreign atoms or ions and lattice defects increase the number of trapping centres and should thus decrease the temperature T_c beyond which the breakdown field F^* shows a falling temperature characteristic. On the other hand, lattice defects and foreign admixtures act as additional scatterer for fast electrons, i.e. they decrease their mean free path. In the low temperature theory this leads to an increase of the breakdown strength. As an example, fused quartz at low temperatures in the range of the positive temperature characteristic for F^* , should have a higher breakdown strength F^* than crystalline quartz; but the transition to the high-temperature range in which F^* has a negative temperature characteristic should start for fused quartz at lower temperatures than for crystalline quartz. Thus at low temperatures fused quartz should have a higher strength than crystalline quartz but this should be reversed at high temperatures. This is well confirmed by experiments by von Hippel & Maurer (1941), whose results are shown in figure 3.

An example of a similar nature is provided by experiments by Austen & Pelzer (1942, 1944) on polythene and on a vinylite. Polythene consists of long hydrocarbon chains $(\text{CH}_2)_n$ and is semicrystalline, as has been shown by Bunn & Alcock (1945). Polyvinyl chloride, on the other hand, is amorphous. It also consists of long chains

$(\text{CH}_2-\text{CHCl})_n$ containing, however, strong CHCl dipoles. These dipoles act as additional scattering centres for fast electrons and decrease their mean free path compared with polythene. Thus at low temperatures the vinylite should have a higher breakdown strength than polythene. In view of its amorphous nature, however, its transition temperature to the falling high temperature characteristic of F^* should be lower than that of polythene. Hence at high temperatures the vinylite should have a lower breakdown strength than polythene. All this has actually been found by Austen & Pelzer.

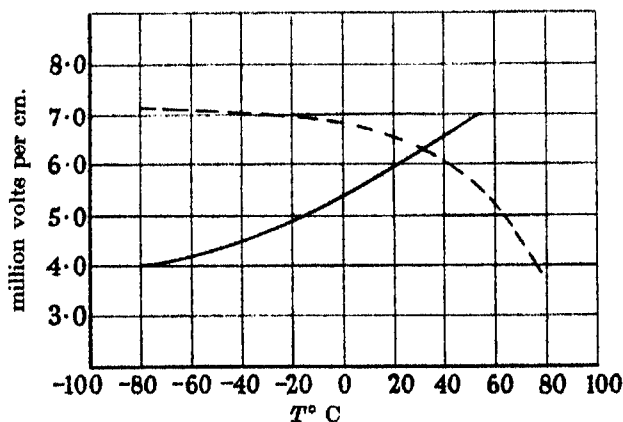


FIGURE 3. Temperature dependence of the breakdown field of crystalline (—) and of fused (---) quartz according to von Hippel & Maurer (1941).

To turn now to a discussion of the temperature dependence (25) of the breakdown field, it seems necessary to find experiments which cover a sufficiently large range. The experiments by von Hippel & Maurer (1941) on soda-lime glass provide such an example. Figure 4 shows that good agreement is obtained if one chooses $\Delta V/k = 3400^\circ$. Other examples are provided by experiments by Hackett & Thomas (1941) on mica, and by Thomas & Griffith (1942) on varnishes. These latter do not, however, cover a sufficiently wide range to lead to a quantitative check. The latter paper contains also measurements of the increase of conductivity with field strength. The F^* -dependence of equation (24) is not entirely in agreement with the experiments which show a slightly slower increase. This may be due to ionic contributions to the conductivity. Equation (24) yields, however, the correct magnitude for the increase in conductivity. Thus for black baking varnish the authors find experimentally $V = 2$ eV, $F^* = 2000$ kV/cm., while from the temperature dependence of F^* , $\Delta V = 0.06$ eV can be deduced. Thus for fields of 500 and 250 kV, equation (25) yields $\log \sigma_{500}/\sigma_{250} = 1$ whereas the experimental value is 2.

From the preceding discussion it follows that the combination of the high- and the low-temperature theory leads to a satisfactory description of breakdown over the whole temperature range. In particular it leads to a simple explanation of the peculiar difference between the high- and the low-temperature behaviour according to which

all influences which tend to increase the breakdown strength in the low-temperature region (e.g. foreign admixtures, increase of temperature) tend to decrease it at high temperatures. Von Hippel (1946) has on several occasions expressed his disbelief in the author's low-temperature theory because in his opinion in fields below breakdown strength not enough fast electrons are available. It seems, however, that the calculations of the accompanying paper prove the existence of a considerable number of fast electrons. Also no other theory can account for all the diverse facts connected with breakdown without using *ad hoc* assumptions.

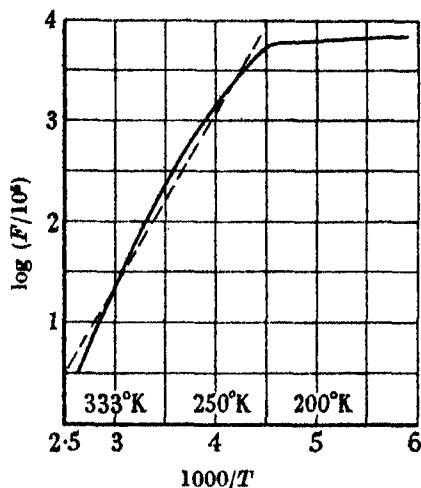


FIGURE 4. Temperature dependence of the dielectric strength of glass. — Experiments by von Hippel & Maurer (1941). --- Theory according to equation (25) with $\Delta V/k = 3400^\circ$.

The present development of the theory is adequate to decide in which cases an improvement of dielectric strength at a given temperature T' (e.g. room temperature) is possible. Clearly the tendency must be to shift the temperature T_c , at which the dielectric strength has its maximum, to T' . If $T_c > T'$ this can be done by foreign admixtures which decrease T_c but increase the breakdown strength. If, on the other hand, $T_c < T'$ it will be required to purify the substance and decrease in other possible ways the number of lattice defects because this shifts T_c to higher temperatures and again increases the breakdown strength.

I am indebted to the British Electrical and Allied Industries Research Association for permission to publish this paper.

Note added on 6 November 1946. I now understand from Mr C. M. Turner [Department of Electrical Engineering, University of Birmingham] that he has measured the field dependence of the conductivity of glass in fields up to about 500 kV/cm., using very short pulses. His results are in very good agreement with equation (24).

REFERENCES

- Austen & Hackett 1939 *Nature*, **143**, 637.
 Austen & Pelzer 1942 *E.R.A. Report* L/T 138.
 Austen & Pelzer 1944 *E.R.A. Report* L/T 149.
 Austen & Whitehead 1940 *Proc. Roy. Soc. A*, **176**, 33.
 Buehl & von Hippel 1939 *Phys. Rev.* **56**, 941.
 Bunn & Alcock 1945 *Trans. Faraday Soc.* **41**, 317.
 Fröhlich 1937 *Proc. Roy. Soc. A*, **160**, 230.
 Fröhlich 1939 *Proc. Roy. Soc. A*, **172**, 94.
 Fröhlich 1941 *Proc. Roy. Soc. A*, **178**, 493.
 Fröhlich 1942 *Phys. Rev.* **61**, 200.
 Fröhlich 1947 *Proc. Roy. Soc. A*, **188**, 532.
 Hackett & Thomas 1941 *J. Instn Elect. Engrs*, **88**, 295.
 von Hippel 1946 *Trans. Faraday Soc.* (in the Press).
 von Hippel & Lee 1941 *Phys. Rev.* **59**, 824.
 von Hippel & Maurer 1941 *Phys. Rev.* **59**, 820.
 Mott & Gurney 1940 *Electronic processes in ionic crystals*. Oxford: Clarendon Press.
 Thomas & Griffith 1942 *J. Instn Elect. Engrs*, **89**, 487.
 Whitehead 1945 *Roy. Coll. Sci. J.* **15**, 105.

Energy distribution and stability of electrons in electric fields

By H. FRÖHLICH, *H. H. Wills Physical Laboratory, University of Bristol*

(Communicated by N. F. Mott, F.R.S.—Received 6 April 1946)

The behaviour of free electrons in ionic crystals in the presence of an external field is studied. It is shown that the usual method of calculating the electric current is incorrect. The correct solution shows that—on the usual assumption that electrons are scattered by the lattice vibrations only—a stationary state is impossible. Stationary conditions can probably be obtained by considering collisions between electrons as well. For very small electron density, however, these latter collisions are negligible. It is shown that in this case the possibility of reaching stationary conditions depends on the behaviour of electrons whose energy is large enough to ionize or excite ions of the lattice.

1. INTRODUCTION AND DISCUSSION

In the present paper it will be shown that no stationary state can be reached on the assumptions conventionally made in calculating the electric current in solids. According to these assumptions, the conduction electrons are considered as free except for their collisions with thermal lattice vibrations and other lattice imperfections. Collisions between electrons are considered as unimportant. While this latter condition can hardly be true for conductors or for good semi-conductors, it should hold for such semi-

conductors for which the density of conduction electrons is so small that they are practically insulators. This case is of importance for the theory of dielectric breakdown, and its implications are discussed in the previous paper. Furthermore, the present paper shows that the mathematical approximation generally made in the calculation of the current is incorrect, although for high electronic density the results obtained can probably be justified if the influence of the mutual collisions between electrons is taken into account. The usual procedure is to assume that on averaging over all directions the perturbations of the energy distribution by a weak external field cancel. This means that the energy exchange between field and electrons, and between electrons and lattice vibrations can be considered as unimportant, a fact which will be disproved in the present paper.

The impossibility of reaching a stationary state in an external field on the assumptions stated above can be demonstrated by an elementary consideration discussed in previous papers (Fröhlich 1937, 1939*a*). Roughly speaking, the rate of energy transfer from the field to an electron is on an average over many collisions proportional to the time of relaxation (approximately the time between two collisions), whereas the rate of energy transfer from the electron to the lattice vibrations is inversely proportional to it. Since for not too slow electrons the time of relaxation increases with the electronic energy, sufficiently fast electrons will on an average gain more energy from the field than they can transfer to the lattice. An electron which by a fluctuation reaches this high-energy region has, therefore, a very small chance of returning to lower energies. Thus all electrons will gradually drift to ever higher energies, i.e. a stationary state is impossible. This elementary consideration, although instructive, is not sufficiently accurate, however, to replace the exact proof given below.

The assumptions leading to this unexpected result cannot be maintained for electronic energies larger than the smallest energy required to excite or ionize an ion or atom of the lattice. This means that an additional investigation will be required in order to find if this new type of collision can establish equilibrium. In the previous elementary treatment it was assumed that equilibrium is always possible unless the external field is so strong that the energy region beyond which an electron will drift to ever higher energies falls below the lowest excitation energy. It has so far not been possible to find an exact proof for this assumption. Nevertheless, the considerations of the present paper establish that the possibility of reaching stationary conditions in an external field is decided by the behaviour of electrons whose energy is of the order of the internal excitation energy, provided the density of the conduction electrons is so small that their mutual collisions can be neglected. This result seems to be of importance in view of recent discussions on this subject (Seeger & Teller 1939; Fröhlich 1939*b*).

The energy distribution (cf. equation (20)) obtained on the assumption that equilibrium can be reached resembles a Maxwell distribution only at small energies, whereas at large energies the distribution function takes on much higher values. The deviation starts at energies at which, roughly speaking, the energy change due to the acceleration by the field between two large-angle collisions, amounts to kT .

2. KINETIC EQUATION AND ENERGY DISTRIBUTION

Consider an ionic crystal with N electrons in the conduction levels. These electrons will be treated as free, i.e. if v is the velocity of such an electron its energy E is given by

$$E = \frac{m}{2} v^2. \quad (1)$$

Later on, instead of E , a dimensionless variable x will be used, defined by

$$x = E/kT, \quad (2)$$

where T is the temperature. An electron will collide with the lattice vibrations, but N is assumed to be so small that mutual collisions between electrons can be neglected. Furthermore, it will be assumed that all lattice vibrations have the same frequency, ν , as is approximately the case in alkali halides.

A distribution function f can now be introduced in such a way that $f \sqrt{x} dx d\Omega$ is the probability of finding at the time t an electron with an energy between x and $x + dx$, and with a direction of motion within a solid angle $d\Omega$ around a given direction. Clearly

$$\int f \sqrt{x} dx d\Omega = N. \quad (3)$$

In equilibrium in the absence of an electric field f will be isotropic, i.e. independent of the direction of the velocity. In the presence of an external field F , however, the equilibrium distribution will depend on the angle θ between F and v but not on the azimuth. Consideration will be given here not only to equilibrium distributions but also to distributions which depend on time t . It will, however, always be assumed that f is independent of the azimuth, i.e. $f = f(x, \theta, t)$. Thus f may be developed in a series of Legendre polynomials

$$f(x, \theta, t) = \sum_{n=0}^{\infty} f_n(x, t) P_n(\cos \theta). \quad (4)$$

This series will be assumed to converge rapidly, and to change only slowly with time. In particular

$$|f_1| \ll f_0 \quad \text{and} \quad \tau \left| \frac{\partial f_1}{\partial t} \right| \ll |f_1|, \quad (5)$$

where τ is the time of relaxation, i.e. roughly the time between two large-angle collisions.

Since $\int P_n d\Omega = 0$ except for $n = 0$, the function f_0 determines the energy distribution of electrons, and $|f_1| \ll f_0$ means that only small deviations from an isotropic distribution will be considered. The function f_1 , on the other hand, determines the electric current I . In fact, considering $P_1 = \cos \theta$, and the orthogonality relations between the P_n , one finds

$$I = e \int v \cos \theta f \sqrt{x} dx d\Omega = \frac{4\pi}{3} e \int v f_1 \sqrt{x} dx. \quad (6)$$

Consider now such temperatures and energies that

$$kT \gg h\nu \quad \text{and} \quad E \gg h\nu. \quad (7)$$

Then in § 3 it will be shown that the functions f_0 and f_1 satisfy the following equations ($f'_0 = \partial f_0 / \partial x$):

$$\frac{\partial f_0}{\partial t} = \frac{B(x)}{kT} \frac{\partial}{\partial x} [(1 + C^2(x))f'_0 + f_0], \quad (8)$$

$$f_1 = -\frac{eF\tau(x)v}{kT} f'_0. \quad (9)$$

Here $B(x)$ is the energy transferred per second from an electron with the energy x to the lattice vibrations. $C^2(x)$ is connected with B and with the quantity

$$A(x) = e^2 F^2 \tau(x) / m \quad (10)$$

by
$$C^2(x) = \frac{2A(x)x}{3B(x)}. \quad (11)$$

$2AN$ represents the rate of energy transfer from the field to the electrons (cf. (21)).

In the following section it will be shown that the energy dependence of τ and of B is given by

$$\tau(x) \propto x^{\frac{1}{2}}, \quad B(x) \propto 1/\sqrt{x}, \quad (12)$$

respectively, provided the energy is larger than an energy E_0 which is of the order of 1 eV. Below E_0 , $\tau \propto x^{\frac{1}{2}}$, but the expression for B remains unaltered as long as (7) is fulfilled. Thus

$$C^2(x) \propto x^3 \quad \text{if} \quad x > E_0/kT \quad \text{and} \quad C^2(x) \propto x^2 \quad \text{if} \quad x < E_0/kT. \quad (13)$$

Equation (8) can be written in the form of a one-dimensional diffusion equation

$$\frac{\partial g}{\partial t} = \frac{\partial j}{\partial x} \quad (14)$$

by introducing in x -space a density function

$$g(x, t) = 4\pi \sqrt{x} f_0(x, t) \quad (15)$$

and a current density

$$j(x, t) = \frac{4\pi \sqrt{x} B(x)}{kT} [(1 + C^2(x))f'_0 + f_0] = \frac{B(x)}{kT} \left[(1 + C^2(x))g' + \frac{2x-1-C^2(x)}{2x} g \right]. \quad (16)$$

In the stationary case, $\partial g / \partial t = 0$, and hence using (14)

$$j(x) = \text{constant}. \quad (17)$$

Now equations (8) or (16) have to satisfy a boundary condition at $x = 0$ which describes the fact that there can be no diffusion of electrons into or from negative-energy regions, $x < 0$. For the stationary case this means that

$$j = 0 \quad \text{for} \quad x = 0. \quad (18)$$

Combining (17) and (18) yields $j(x) = 0$, or using (16)*

$$(1 + C^2(x))f'_0 + f_0 = 0. \quad (19)$$

Hence the energy distribution in the stationary case must have the form

$$f_0(x) = a \exp \left\{ - \int_0^x \frac{d\eta}{1 + C^2(\eta)} \right\}, \quad (20)$$

where a is constant. For small energies, $C^2 \ll 1$, and hence $f_0(x) \simeq ae^{-x}$ becomes a Maxwell distribution. For large x , however, since C^2 increases $\propto x^3$ (cf. (13)), the value $C^2 = 1$ will be reached (when $eF\tau(x)v \simeq kT$) and surpassed, however weak the external field (unless $F = 0$). In this region f_0/a deviates from a Maxwell distribution and asymptotically approaches the finite value

$$\frac{f_0}{a} = \exp \left\{ - \int_0^\infty \frac{d\eta}{1 + C^2(\eta)} \right\} \neq 0 \quad \text{if } x \rightarrow \infty.$$

This means that the normalization integral (3) diverges, i.e. (20) would require the presence of an infinite number of electrons. Thus it must be concluded that a stationary solution is impossible. It should be emphasized that this result rests on the fact that f_0/a remains finite as $x \rightarrow \infty$. Thus if beyond a certain $x = x'$ the above consideration becomes invalid (e.g. because the electrons can ionize or excite ions of the lattice) a stationary state may be possible and (20) holds up to $x = x'$. This case would require further investigation. So far, however, I have not been able to incorporate internal ionization processes in a rigid way into the calculation of the distribution function.

It is worth noticing that the energy-distribution function $g(x)$, (15), has a minimum at $x = x_m$ given by (using (20), (11) and $x_m \gg 1$)

$$A(x_m) = 3B(x_m).$$

On the other hand, from elementary considerations it has been proposed previously (Fröhlich 1937) that stationary conditions should no longer be possible if the energy x'' for which $A(x) = B(x)$ is smaller than the energy x' required for internal ionization. In the light of the present investigation, disregarding the factor 3 in the above formula, this means that the distribution function $g(x)$ should decrease with increasing x at least up to $x = x'$, i.e. throughout the range in which the present theory can claim validity.

* The postulation of the boundary conditions (18) requires some further justification because all the formulae are valid only if $x \gg \hbar\nu/kT$ (7). This justification can be provided because in the neighbourhood of $x = \hbar\nu/kT$, $C^2(x)$ becomes negligibly small even for the strongest fields contemplated in this paper. Thus even if the quantitative expression for $C^2(x)$ is incorrect near $x = 0$, equation (18) leads to

$$f'_0 + f_0 = 0 \quad \text{for } x = 0$$

as long as $C^2(0) \ll 1$. But this equation can easily be derived directly for the case $x < \hbar\nu/kT$ using the methods of § 3 and considering that an electron with an energy $E < \hbar\nu$ can absorb but not emit a quantum $\hbar\nu$.

Furthermore, it can be shown that stationary conditions are impossible even if electrons are allowed to enter or leave at $x = 0$, i.e. if the boundary condition (18) is given up. A necessary condition for a stationary solution of (8) is $f_0 = 0$ for $x \rightarrow \infty$. Then

$$f_0 = a \exp \left\{ - \int_0^x \frac{d\eta}{1 + C^2(\eta)} \right\} - a \exp \left\{ - \int_0^\infty \frac{d\eta}{1 + C^2(\eta)} \right\}.$$

In contrast to (20) this function can be normalized. The current I , however, if calculated from (6) (using (9), (12) and the above value for f_0) is found to diverge because of the very large contribution of the fast electrons.* Again one must conclude that a different behaviour of the fast electrons is required to establish stationary conditions. This means that, irrespective of the boundary condition, no stationary solution of (8) exists which is always positive and which leads to finite values for N and I .

Finally it should be of interest to consider a simple non-stationary case. Suppose at the time $t = 0$ the distribution function to be very narrow, e.g.

$$4\pi f_0(x) = \delta(x - x_0),$$

where $\delta(x - x_0)$ is big near x_0 only and satisfies (of. (3))

$$N = \int \delta(x - x_0) \sqrt{x} dx = \sqrt{x_0} \int \delta(x - x_0) dx.$$

Using (8) and (2) the average rate of change \dot{U} of the energy is found to be

$$\dot{U} = \frac{4\pi}{N} kT \int_0^\infty x \frac{\partial f_0}{\partial t} \sqrt{x} dx = \frac{4\pi}{N} \int_0^\infty x^{\frac{1}{2}} B(x) \frac{\partial}{\partial x} [(1 + C^2)f'_0 + f_0] dx.$$

According to (12), $B(x)\sqrt{x}$ is independent of x . Then, integrating several times by parts one finds, using $C^2 \propto x^2$, near $x = x_0$

$$\dot{U} = -B(x_0) + 2A(x_0).$$

Making use of (6) and (9) yields $2A = IF/N$, where I is the current carried by the electrons. Thus

$$\dot{U} = -B(x_0) + IF/N, \quad (21)$$

a result which also follows from elementary considerations. In particular, since A increases and B decreases with x , \dot{U} is positive for large x but negative for small x , a fact which has been used in § 1 to demonstrate the impossibility of reaching stationary conditions.

3. DERIVATION OF THE KINETIC EQUATION

To derive equations (8) and (9), Bloch's (1928) theory of conductivity will be adapted to the case treated in the present paper. The series (4) will be used to represent the distribution function f , and it will be considered sufficient to use the f_0 and

* Nevertheless $|f_1| \ll f_0$ is fulfilled.

f_1 terms only* as indicated by (5). Bloch assumed f_0 to be the distribution in the absence of an external field. This is not permissible because f_0 has to be calculated from the theory and cannot be assumed *a priori*.†

Following Bloch let the rate of change of f be decomposed into the two components f_F and f_C , denoting the contributions of the field and of the collisions respectively, i.e.

$$\frac{\partial f}{\partial t} = f_F + f_C, \quad (22)$$

where assuming the field to be in the z -direction ($v_z = z$ -component of \mathbf{v})

$$f_F = -\frac{eF}{m} \frac{\partial f}{\partial v_z} = -\frac{eFv}{kT} \left(\frac{1}{3x} \frac{\partial}{\partial x} (f_1 x) + \frac{\partial f_0}{\partial x} \cos \theta \right) + \text{terms with } P_2 + \dots \quad (23)$$

To calculate f_C use will be made of the calculation of the scattering of electrons by the lattice vibrations of a polar lattice of the NaCl type carried out in a previous paper‡ (Fröhlich 1937, quoted as 'A'). Let \mathbf{w} be the wave vector of a polarization wave, and assume that all these waves have the same frequency ν . Then an electron can either absorb or emit a quantum $\hbar\nu$ and hereby change its own velocity \mathbf{v} in such a way that the momentum law is fulfilled. According to 'A' the transition probability for absorption of an $\hbar\nu$ connected with the transition of the velocity from \mathbf{v} to \mathbf{v}' is given by

$$\phi_{\mathbf{v}, \mathbf{v}'}^a = \left(\frac{2\pi e^2}{a^3} \right)^2 \frac{1}{2M\hbar\nu N'} \frac{n}{w^2} \frac{\partial}{\partial t} \left(\frac{\sin \xi t}{\xi} \right)^2, \quad (24)$$

where the momentum law requires

$$\mathbf{v}' = \mathbf{v} + \frac{\hbar}{m} \mathbf{w}. \quad (25)$$

Here the charge of each of the $2N'$ ions is supposed to be $\pm e$, M is their resultant mass, a is the lattice distance and n the number of quanta with wave number \mathbf{w} present before the transition takes place. Furthermore,

$$2\hbar\xi = E' - E - \hbar\nu. \quad (26)$$

In thermal equilibrium

$$n = \frac{1}{e^{\hbar\nu/kT} - 1}, \quad 1 + n = e^{\hbar\nu/kT} n. \quad (27)$$

* Actually the solution of the stationary equation (19) satisfies $|f_1| \ll |f_0|$ as can be seen by inserting f'_0 from (19) into (9) because, considering (7), $B\tau/kT \simeq \hbar\nu/kT \ll 1$.

† The present method has been used previously by Landau & Kompanejev (1934). Unfortunately, these authors use a wrong energy dependence for a transition probability which they denote by W . They quote Bloch as giving $W \propto \sqrt{E}$, whereas actually $W \propto 1/\sqrt{E}$. This mistake completely distorts the results.

‡ The following errors should be corrected in 'A':

(1) $(2N)^\dagger$ must be replaced by N^\dagger in equation (4a), (10) and in the equation preceding (11a).

(2) $\hbar\xi$ on p. 236 must be replaced by $2\hbar\xi$.

(3) The right-hand sides of (11a) and (12a) must be multiplied by $1/N$.

The probability for the inverse process (emission of $h\nu$, $\mathbf{v}' \rightarrow \mathbf{v}$) is given by

$$\phi_{\mathbf{v}', \mathbf{v}}^0 = \phi_{\mathbf{v}, \mathbf{v}'}^0 \frac{1+n}{n}. \quad (28)$$

Furthermore, the transition probabilities for emission of an $h\nu$ connected with an electronic transition $\mathbf{v} \rightarrow \mathbf{v}''$ and for the inverse process will be required. They can be obtained from (24) and (28) by replacing \mathbf{v}' by \mathbf{v} and \mathbf{v} by \mathbf{v}'' , where

$$\mathbf{v}'' = \mathbf{v} - \frac{\hbar}{m} \mathbf{w}. \quad (29)$$

To find the total rate of change $f_C(x, \theta)$ of the distribution function for the velocity v (v is determined by x, θ and an azimuth) all the transitions from and into the state v have to be considered, i.e. $\mathbf{v} \rightarrow \mathbf{v}'$, $\mathbf{v} \rightarrow \mathbf{v}''$, $\mathbf{v}' \rightarrow \mathbf{v}$, $\mathbf{v}'' \rightarrow \mathbf{v}$. They are given by the products of the respective transition probabilities ϕ with the distribution function in the initial state, i.e. with $f(x, \theta)$, $f(x, \theta)$, $f(x', \theta')$, $f(x'', \theta'')$ respectively.* Finally, the contributions of all possible wave numbers have to be added. Their number per unit volume in w -space is given by $2a^3 N' \sin \alpha d\alpha d\psi w^2 dw / (2\pi)^3$ if polar-coordinates in w -space are used with \mathbf{v}/v as the direction of the axis, i.e. with α as angle between \mathbf{v} and \mathbf{w} . ψ is the azimuth around v . Then

$$f_C(x, \theta, t) = \frac{2a^3 N'}{(2\pi)^3} \left(\frac{2\pi e^2}{a^3} \right)^2 \frac{n}{2M\hbar\nu N'} (J_1 - J_2), \quad (30)$$

where J_1 contains the contributions due to the transitions $\mathbf{v} \rightarrow \mathbf{v}'$ and $\mathbf{v}' \rightarrow \mathbf{v}$, while J_2 refers to $\mathbf{v} \rightarrow \mathbf{v}''$, $\mathbf{v}'' \rightarrow \mathbf{v}$. Thus

$$J_1 = \int [-f(x, \theta) + e^{h\nu/kT} f(x', \theta')] \frac{\partial}{\partial t} \left(\frac{\sin \xi t}{\xi} \right)^2 dw \sin \alpha d\alpha d\psi. \quad (31)$$

J_2 is obtained from this expression by replacing \mathbf{v}' by \mathbf{v} and \mathbf{v} by \mathbf{v}'' . To carry out the α -integration in (31) use will be made of equation (25) which leads to

$$v'^2 = v^2 + \left(\frac{\hbar w}{m} \right)^2 - 2v \frac{\hbar w}{m} \cos \alpha.$$

Hence with the help of (26) and (1)

$$\sin \alpha d\alpha = -d(\cos \alpha) = \frac{m}{2\hbar} \frac{d(v'^2)}{vw} = \frac{2d\xi}{vw}.$$

Since $\left(\frac{\sin \xi t}{\xi} \right)^2$ has a steep maximum at $\xi = 0$, (31) becomes

$$J_1 = \frac{2\pi}{v} \int [-f(x, \theta) + e^{h\nu/kT} f(x'_0, \theta'_0)] \frac{dw}{w} d\psi, \quad (32)$$

* The electronic density is considered to be so small that the Pauli principle becomes irrelevant because the final state is nearly always unoccupied.

where x'_0 and θ'_0 are the values of x' and θ' if $\xi = 0$. Thus from (26), $x'_0 = x_0 + h\nu/kT$. θ'_0 is given by

$$\cos \theta'_0 = \cos \theta \cos \beta + \sin \theta \sin \beta \cos \psi, \quad (33)$$

where β is the angle between v and v' on the condition that $\xi = 0$. According to (26) and (1) this means that $v'^2 = v^2 + 2h\nu/m$. With the use of $E \gg h\nu$ (7), and of eqn. (25), $\cos \beta$ is thus found to be given by

$$\cos \beta = \frac{v^2 + v'^2 - \hbar^2 w^2/m}{2vv'} \simeq 1 - \frac{w^2 \hbar^2}{2v^2 m^2}. \quad (34)$$

According to (4) introduce now $f(x, \theta) = f_0(x) + f_1(x) \cos \theta$ and integrate (32) with respect to ψ . Then the terms proportional to $\sin \theta \sin \beta$ vanish and thus

$$J_1 = \frac{4\pi^2}{v} \int \left\{ [-f_0(x) + f_0(x + h\nu/kT) e^{h\nu/kT}] + \left[-f_1(x) + f_1(x + h\nu/kT) e^{h\nu/kT} \left(1 - \frac{\hbar^2 w^2}{2m^2 v^2} \right) \right] \cos \theta \right\} \frac{dw}{w}. \quad (35)$$

J_2 is now easily obtained by making the replacements mentioned above. For the x and θ 's this means $x \rightarrow x - h\nu/kT$, $x + h\nu/kT \rightarrow x$, $\theta \rightarrow \theta'$, $\theta' \rightarrow \theta$. θ' is given by a formula similar to (33). In view of the approximation $h\nu \ll E$, $\cos \beta$ has now the same value (34) as before. Thus

$$J_2 = \frac{4\pi^2}{v} \int \left\{ [-f_0(x - h\nu/kT) + f_0(x) e^{h\nu/kT}] + \left[-f_1(x - h\nu/kT) \left(1 - \frac{\hbar^2 w^2}{2m^2 v^2} \right) + f_1(x) e^{h\nu/kT} \right] \cos \theta \right\} \frac{dw}{w}. \quad (36)$$

Considering (7) $f(x \pm h\nu/kT)$ and $e^{h\nu/kT}$ can now be developed into a power series in $h\nu/kT$ which converges rapidly. Then keeping in $(J_1 - J_2)$ the lowest non-vanishing term equation (30) yields (use $n \sim kT/h\nu$)

$$f_C(x, \theta, t) = \frac{4\pi^2}{v} \frac{e^4}{2\pi M a^3} \frac{kT}{(h\nu)^2} \left\{ \left(\frac{h\nu}{kT} \right)^2 \frac{\partial}{\partial x} (f'_0 + f_0) \int \frac{dw}{w} - \frac{\hbar^2}{m^2 v^2} f_1 \cos \theta \int w dw \right\}. \quad (37)$$

The quantities $B(x)$ and $\tau(x)$ introduced in equations (8) and (9) will now be defined as

$$B(x) = \frac{2\pi e^4}{M a^3} \frac{1}{v} \int \frac{dw}{w}, \quad \frac{1}{\tau(x)} = \frac{e^4 kT}{2\pi m^2 M a^3 v^2} \frac{1}{v^3} \int w dw. \quad (38)$$

The integrals over w can easily be evaluated. Their limits have been discussed in 'A', and the result for B and τ agrees with that given in 'A', equations (16) and (17).

Making use of (1) it is seen that equation (12) is fulfilled if $\int dw/w = \log \gamma$ [cf. 'A', eqn. (17)] is considered to be independent of E . According to 'A' this represents a good approximation.

Finally, introducing (37), (38) and (23) into (22) leads to the kinetic equations (8) and (9).

REFERENCES

- Bloch 1928 *Z. Phys.* **52**, 555.
 Fröhlich 1937 *Proc. Roy. Soc. A*, **160**, 230.
 Fröhlich 1939a *Rep. Progr. Phys.* **6**, 411.
 Fröhlich 1939b *Phys. Rev.* **56**, 349.
 Landau & Kompanejev 1934 *Z. Phys. Sowjet.* **6**, 163.
 Seeger & Teller 1939 *Phys. Rev.* **56**, 352.

Intensities of γ rays, studied by means of their Compton secondaries

BY K. SIEGBAHN, *Nobel Institute for Physics, Stockholm*

(Communicated by Sir Owen Richardson, F.R.S.—Received 10 April 1946)

Methods for investigating the γ rays from active isotopes are discussed. A method is described which permits the study of very faint samples. It is possible by this method to obtain information about the intensities as well as the energies of the γ rays. Such investigations necessitate knowledge of the efficiency curve for a secondary-electron radiator. With the aid of the known term schemes for Na^{24} , Mn^{56} and Cl^{38} an efficiency curve for copper is constructed.

INTRODUCTION

In order to be able to establish a complete term scheme for a radioactive disintegration, it is usually necessary to determine the energies as well as the intensities of the β and γ components of the radiation. If the β spectrum has several components, these may often be separated by constructing the corresponding Fermi diagram. Such a Fermi analysis will, however, meet with great difficulties, if the upper limits of the different β components lie close together or if the components of lower energy have much lower intensities than the others. It is, in such cases, often necessary to determine in some way the difference in energy between the different β components, which is obtained, as is well known, in the form of γ radiation.

It sometimes occurs that the β transition from the active isotope direct to the ground state of the accompanying isotope is so strictly forbidden that it is absent. The ground state is in this case attained by the emission of one or several γ quanta after the β spectrum. Several of the lowest excited levels in the final nucleus may thus be missed, if the γ radiation in the disintegration is not studied.

Likewise, the probability of transitions between the excited levels may often be determined by the study of the relation between the β intensities; intensity investigations on the γ radiation must sometimes also be resorted to.

METHODS FOR INVESTIGATING γ RADIATION

A number of methods are available for such investigations. The absorption method is probably that hitherto most employed. Its applicability is, however, limited to the occasions when there are only one or a few γ lines. The method yields good results only if very favourable geometrical conditions are provided (large distance between sample and Geiger-Müller tube, well-canalized radiation path, etc.), together with protection against dispersed radiation. Several γ lines, occurring simultaneously, can be satisfactorily separated only if their energies differ considerably.

The other methods tried include the application of the nuclear photo-effect in D and Be, when the energy of the neutrons generated is studied. Only γ radiation with higher energy than the threshold value for the reaction in question (2.18 and 1.62 MeV respectively) can be investigated in this way (Goldhaber, Klaiber & Scharff-Goldhaber 1944). The pair formation for γ radiation $> 2 mc^2$ may also be used in γ energy determinations, though this method is rather laborious (Krugel & Ogle 1945). Coincidence investigations (β - γ and γ - γ coincidences) frequently yield valuable information, provided that the efficiency curves of the β and γ tubes as functions of the energy are well known. The last-named method is especially suited to complement the methods for the study of γ radiation, which will be reviewed in the following account.

THE PHOTO-COMPTON METHOD

The most successful method at present for the investigation of γ radiation is the study of the secondary electrons emitted by the photo- and Compton effects in a suitable secondary radiator. A frequent procedure is to place a thin lead or aluminium foil (~ 0.1 mm.) in a Wilson chamber at so large a distance from the sample that the direction of the γ radiation relative to the emitted photo- or Compton electrons can be regarded as known within certain limits (Richardson & Kurie 1936). One advantage of this arrangement is that no mixing need occur between photo- and Compton electrons on the one hand and β particles from the continuous spectrum of the active isotope on the other, since only the electrons emanating from the foil are measured. The disadvantage of the method, however, is that it has a comparatively low 'light intensity', owing to the large distance between the sample and the secondary radiator.

If the sample is placed in contact with the secondary radiator, this has to be so thick that the continuous β spectrum is absorbed in the radiator. Combined with the β spectrograph, this technique has been employed by Curran, Dee & Strothers (1940), Mandeville (1942), Deutsch, Elliott & Evans (1944), and the present author. Curran *et al.* and Mandeville have used largely the same technique: the sample is placed behind an obliquely placed, extended aluminium plate, which serves as a Compton radiator in a semicircular spectrograph. The secondary electrons are registered with the aid of the three Geiger-Müller tubes placed in a row, joined together in coincidence-coupling. The spectrograph is directly connected with the Geiger-Müller tubes so that the same pressure prevails throughout (~ 100 mm.). The lowest γ energies capable of investigation are reported by Mandeville to be 0.5 MeV. Each γ energy gives, according

to Mandeville, a typically peaked, symmetrical Compton distribution. The intensities of the different γ components are obtained by dividing the areas of the Compton distributions by the corresponding Compton coefficients and the maximum ranges of the Compton electrons.

An essentially different technique has been developed by Deutsch *et al.* (1944) and the present author. The sample is pressed into a small, cylindrical radiator ($\phi = 8$ mm.), whose walls are thick enough to absorb the continuous spectrum. Copper is a suitable material to limit the size of the radiator, for it has a high density but an atomic number low enough to give almost exclusively Compton effect at energies > 0.5 MeV. If the outside of this radiator is covered with a thin lead foil, photo-electrons from the thin lead foil are obtained in addition to the Compton electrons. If the secondary-electron distribution, obtained in experiments with and without lead foil, is subtracted, there remain the photo-lines from lead. Very accurate energy determinations of the γ radiation may be attained in this way, as the peak of a line can be more accurately measured than the sloping edge of a Compton distribution.

The radiation from the secondary radiator is studied in a lens β spectrograph of high light intensity (in course of publication). It has proved possible, by adjusting slits and diaphragm to a resolving power of 5 % (defined from the half-breadth value of a β line), to measure accurately the energy as well as the intensity for samples whose γ radiations are no higher than $\sim 1 \mu\text{C}$ Ra- γ equivalence/ γ line. The Geiger-Müller tube in the β lens spectrograph will, for obvious reasons, be so far from the sample that the γ radiation from the latter will not considerably influence the zero effect of the tube. A special arrangement for coincidence-coupled β tubes is also unnecessary. It is difficult to fix a limit for the lowest γ energies susceptible of study by this method. This depends mainly on the thickness of the copper radiator necessary for the absorption of the β spectrum. Hitherto, γ energies of < 100 keV (Slätis 1946) have been successfully investigated. Apart from the γ absorption in the radiator, however, there is nothing to prevent the study of still lower energies.

The photo- and Compton methods seem to complement each other excellently. The photo-method is superior for γ energies of, for example, 800 keV and below, owing to the fact that the photo-lines obtained within this range of energy lie so high above the corresponding Compton distributions that they can easily be subtracted from the total curve.

As the energy becomes higher, however, the Compton method will be more and more applicable. It is, of course, always possible to separate satisfactorily the photo-line from the corresponding Compton edge, if the resolving power of the spectrograph is sufficiently high (which involves a reduced intensity). It proves difficult, however, at the resolving power used by the present author (5 %), to obtain clearly observable photo-lines for large γ energies. These often reveal themselves as an inconsiderable 'terracing' of the Compton edge. Obviously the intensity of such a photo-line is particularly uncertain. This does not, however, apply to the Compton distribution. The continuous-energy distribution of the latter makes it possible to base the intensity measurements on a number of points of measurement, all of equal value. The rather

characteristic form that the Compton distributions are proved to possess by closer experimental study further increases the possibility of obtaining good γ intensity determinations based on this effect.

The problem of how to calculate, from photo-lines, the corresponding γ line intensities, has earlier been dealt with in a convincing manner by Deutsch *et al.* (1944) and will therefore not be discussed here. The remaining part of this paper will instead deal with determinations of the γ intensity from the corresponding Compton distributions.

THE EFFICIENCY CURVE FOR A COPPER RADIATOR

The number of Compton electrons (i.e. the area below the Compton distribution) emanating from the radiator and registered in the spectrograph, is a function of the energy of the γ line. This is due to the facts that the Compton coefficient varies with the energy and that the absorption of the liberated secondary electrons also varies with the energy. These two effects act in opposing senses, since the Compton absorption coefficient decreases as the energy increases, whereas the range of the secondary electrons naturally increases as the β energy increases. This latter functional process is steeper than the former, which implies that the 'efficiency' of the radiator for different γ energies is an ascending function. It is evidently necessary to know this function, in order to calculate the intensity of the γ line from the known area of the Compton distribution.

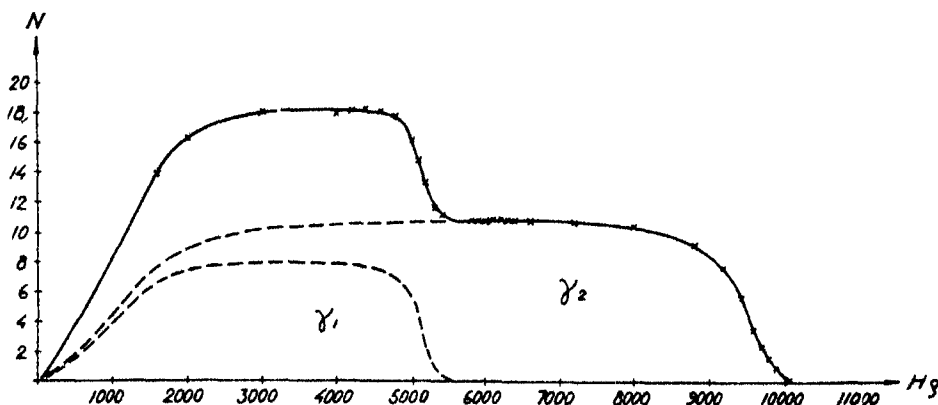


FIGURE 1. γ spectrum of Na^{24} .

Possible though it be to treat this problem theoretically, the following empirical procedure is certainly simpler and possibly more reliable. The relations are studied between the areas under the Compton distributions originating from different γ lines, the relative intensities of which may be considered as known from other data in the term scheme of the disintegration.

At present, unfortunately, but few cases are known of disintegrations where the term scheme is completely clear. These include a still lower number for which the intensity data are also determined. As a basis for the efficiency curve for copper as

a secondary radiator, the disintegrations of the active isotopes Na^{24} , Mn^{56} and Cl^{38} have been chosen.

Na^{24} has a simple β spectrum, and the two γ lines (1.38 and 2.76 MeV) must be emitted in cascade after the β spectrum (Elliot, Deutsch & Roberts 1943; Siegbahn 1946*b*). Accordingly, the intensities of the two γ lines are in the proportion 1 : 1. Figure 1 shows the appearance of the Compton secondaries from copper for these two γ lines. The curve may be resolved into two components with largely similar form. It is intriguing to note that each curve is horizontal over a large energy range in the central part. The ratio of the areas is 2.65.

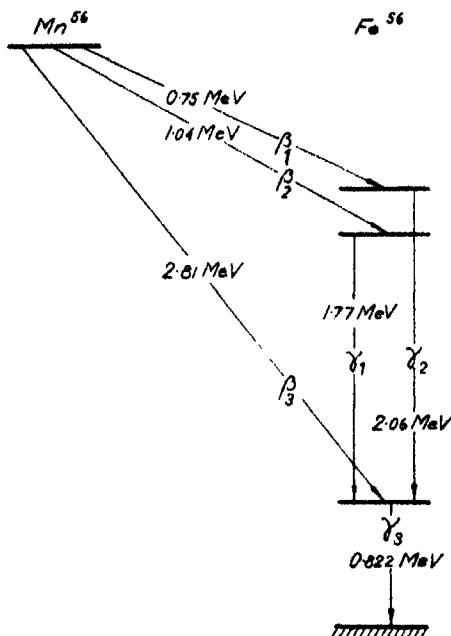


FIGURE 2. Term scheme for the disintegration of Mn^{56} .

The disintegration of Mn^{56} is considerably more complicated (Siegbahn 1946*a*). It is illustrated by the term scheme in figure 2. Knowing the intensities of the various β transitions, one may calculate the corresponding data for the γ transitions. The ratio of I_{β_1} and I_{β_2} is a little uncertain (approx. 2 : 3), whereas $(I_{\beta_1} + I_{\beta_2}) : I_{\beta_3}$ may be given with greater accuracy as 1.00. According to the term scheme, the sum of I_{γ_1} ($E_{\gamma_1} = 1.77$ MeV) and I_{γ_2} ($E_{\gamma_2} = 2.06$ MeV) must then be 50 % of I_{γ_3} ($E_{\gamma_3} = 0.822$ MeV). The mean of the energies of γ_1 and γ_2 may then reasonably be placed at 1.87 MeV, considering that γ_1 is stronger than γ_2 . Figure 3 shows the recording of the Compton secondaries from copper for the γ radiation. The different components have, according to the figure, approximately the same characteristic forms as in the case of Na^{24} . The ratio of the sum of Compton distributions for γ_1 and γ_2 to that of γ_3 is 1.36. Thus, the efficiency of the copper radiator, considering that $I_{\gamma_1} + I_{\gamma_2} = 50$ % of I_{γ_3} , is 2.72 times greater at 1.87 MeV than at 0.822 MeV.

Cl^{38} also has a rather complex disintegration, which is clear from the term scheme in figure 4 (Hole & Siegbahn 1946). The proportions of the intensities of β_1 , β_2 and β_3 are (from the Fermi analysis of the β spectrum) 36:11:53. Hence, the conclusion may be drawn that the ratio of I_{γ_1} (1.60 MeV) and I_{γ_2} (2.15 MeV) should be 43:57.

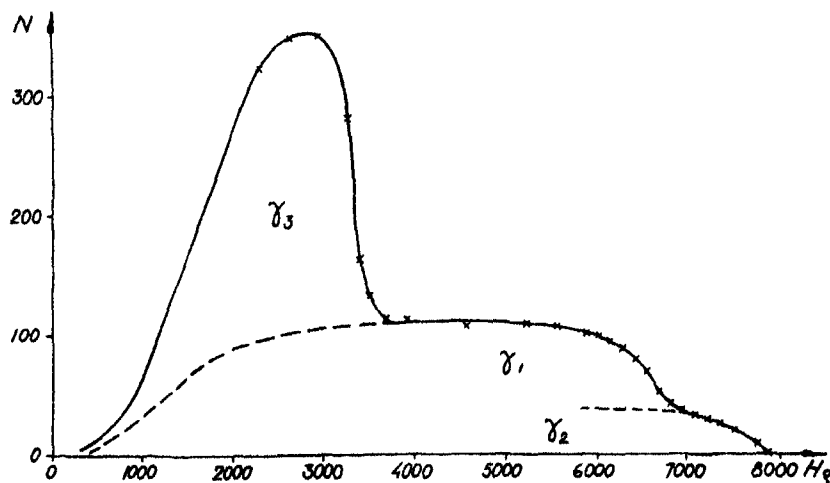


FIGURE 3. γ spectrum of Mn^{56} .

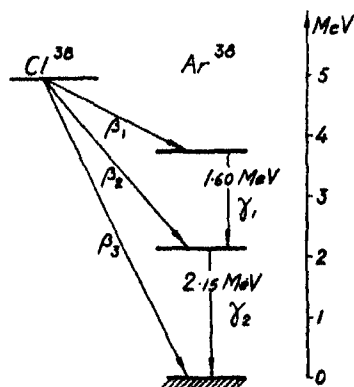


FIGURE 4. Term scheme for the disintegration of Cl^{38} .

The Compton recording of the two γ components is shown in figure 5, where the characteristic energy distribution of the secondary electrons is again to be seen. The ratio of the areas of the two components is 60:120. The efficiency values of the copper radiator at 2.12 and 1.65 MeV are accordingly in the ratio $120/57:60/43 = 1.50$.

The efficiency curve of the copper radiator has been constructed in figure 6 from the above results. This has been done in the following way. The efficiency at the energy 1.38 MeV has been arbitrarily taken as unity. An additional point of the curve at 2.76 MeV may at once be obtained from Na^{24} . Further, the ratio of the efficiency values

at 1.85 and 0.822 MeV (from Mn^{56}) is known to be 2.72. Since the curve may be supposed to have an approximately monotonic course, there is no difficulty in adjusting the two points from Mn^{56} so that an even curve may be drawn through all four points.

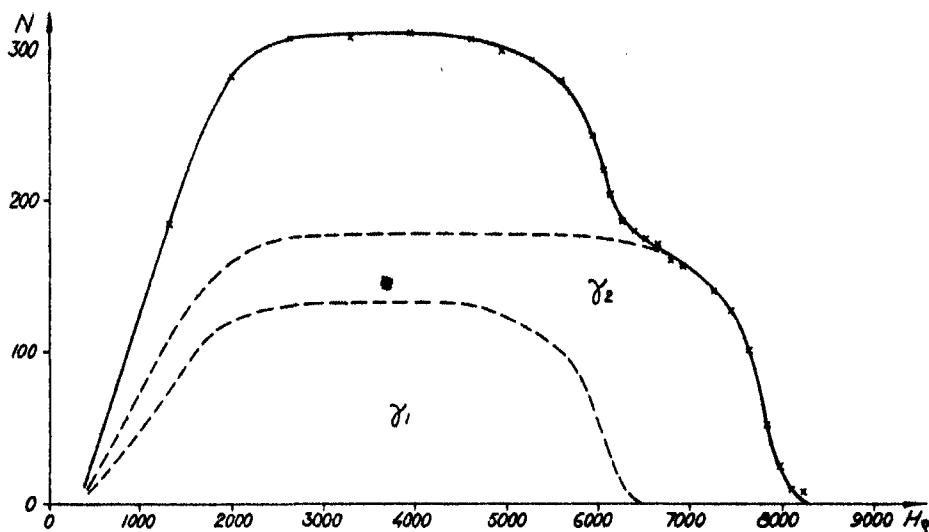


FIGURE 5. γ spectrum of Cl^{38} .

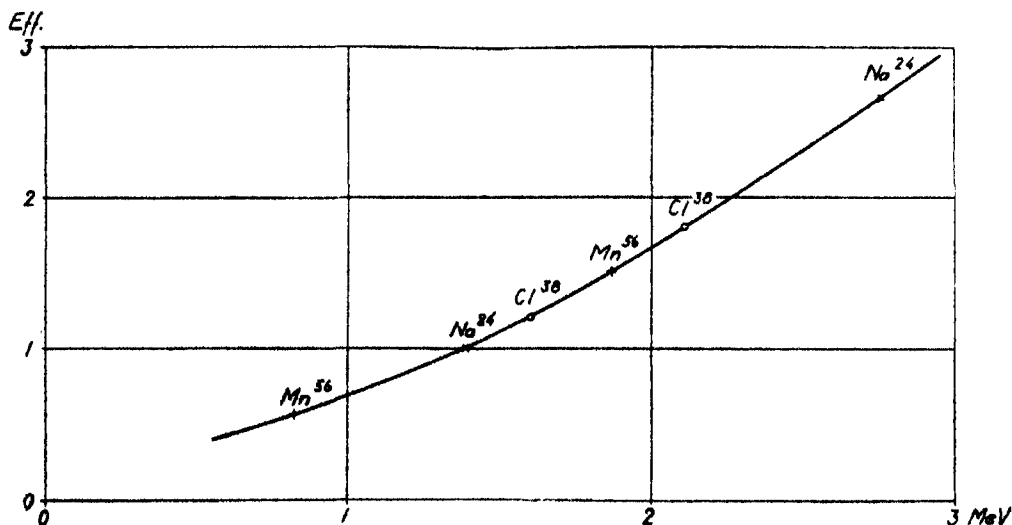


FIGURE 6. The efficiency curve for a Cu radiator.

A very satisfactory check on the efficiency curve thus obtained is provided by the fact that the two points obtained from Cl^{38} also lie on the curve.

The accuracy of intensity determinations based on this method may be estimated as 10%, possibly somewhat higher. About the same value is given for the photo-method

by Deutsch *et al.* (1944). The largest error lies in the sometimes rather difficult separation of the γ components. It is, however, helpful to know that each component seems generally to have the characteristic form shown by the above examples. The geometrical form of the radiator does probably not affect, to any considerable extent, this energy distribution, although this matter calls for somewhat closer study. Nevertheless, the efficiency of the radiator as a function of the energy appears to be inappreciably influenced by small changes in the form of the energy distributions of the components.

REFERENCES

- Curran, S. C., Dee, P. I. & Strothers, J. E. 1940 *Proc. Roy. Soc. A*, **174**, 546.
Deutsch, M., Elliott, L. G. & Evans, R. D. 1944 *Rev. Sci. Instrum.* **15**, 178.
Elliott, L. G., Deutsch, M. & Roberts, A. 1943 *Phys. Rev.* **63**, 386.
Goldhaber, M., Klaiber, G. S. & Scharff-Goldhaber, G. 1944 *Phys. Rev.* **65**, 61.
Hole, N. & Siegbahn, K. 1946 *Ark. Mat. Astr. Fys.* **33A**, no. 9.
Krugel, P. G. & Ogle, W. E. 1945 *Phys. Rev.* **67**, 273.
Mandeville, C. E. 1942 *Phys. Rev.* **62**, 309.
Richardson, J. R. & Kurie, F. N. D. 1936 *Phys. Rev.* **50**, 999.
Siegbahn, K. 1946*a* *Ark. Mat. Astr. Fys.* **33A**, no. 10.
Siegbahn, K. 1946*b* *Phys. Rev.* (in the Press).
Slätis, H. 1946 *Ark. Mat. Astr. Fys.* **33A**, no. 17.

INDEX TO VOLUME 188 (A)

- Accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis (Booth), 77.
- Aerodynamic drag of the earth's surface and the value of von Karman's constant in the lower atmosphere (Sheppard), 208.
- Aerofoils, hydrodynamical theory of moving (Morris), 439.
- Anisotropy of thermal expansion as a cause of deformation in metals and alloys (Boas & Honeycombe), 427.
- Anniversary Address (Robinson), 143.
- Archer, E. M. The crystal structure of meta-dinitrobenzene, 51.
- Atomic co-ordinates, accuracy (Booth), 77.
- Bakerian Lecture. The more recent work on the reaction between hydrogen and oxygen (Hinshelwood), 1.
- Bates, D. R. The quantal calculation of the photo-ionization cross-section of atomic potassium, 350.
- Begbie, G. H. Thermal scattering of X-rays by crystals. II. The thermal scattering of the face-centred cubic and the close-packed hexagonal lattices, 189.
- Begbie, G. H. & Born, M. Thermal scattering of X-rays in crystals. I. Dynamical foundation, 179.
- Blears, J. Measurement of the ultimate pressures of oil-diffusion pumps, 62.
- Bloom, H. & Heymann, E. The electric conductivity and the activation energy of ionic migration of molten salts and their mixtures, 392.
- Boas, W. & Honeycombe, R. W. K. The anisotropy of thermal expansion as a cause of deformation in metals and alloys, 427.
- Booth, A. D. The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis, 77.
- Born, M. *See* Begbie & Born.
- Born, M. & Bradburn, M. The theory of the Raman effect in crystals in particular rock-salt, 161.
- Born, M. & Green, H. S. A general kinetic theory of liquids. 1. The molecular distribution functions, 10.
- Bowden, F. P., Mulcahy, M. F. R., Vines, R. G. & Yoffe, A. The detonation of liquid explosives by gentle impact. The effect of minute gas spaces, 291.
- Bowden, F. P., Mulcahy, M. F. R., Vines, R. G. & Yoffe, A. The period of impact, the time of initiation and the rate of growth of the explosion of nitroglycerine, 311.
- Bowden, F. P., Stone, M. A. & Tudor, G. K. Hot spots on rubbing surfaces and the detonation of explosives by friction, 329.
- Bradburn, M. *See* Born & Bradburn.
- Brummage, K. G. An electron diffraction study of the structure of thin films of normal paraffins, 414.
- Corner, J. The internal ballistics of a gun after shot ejection, 255.
- Corner, J. The internal ballistics of a leaking gun, 237.
- Corrosion fatigue of mild steel (Evans & Simnad), 372.
- Crystal structure of meta-dinitrobenzene (Archer), 51.
- Crystal structure of some molecular complexes of 4:4'-dinitrodiphenyl. I. The complex with 4-hydroxydiphenyl (Saunders), 31.

- Detonation of liquid explosives by gentle impact. The effect of minute gas spaces (Bowden, Mulcahy, Vines & Yoffe), 291.
- Effect of temperature on the intensity of X-ray reflexion (Owen & Williams), 509.
- Electric conductivity and the activation energy of ionic migration of molten salts and their mixtures (Bloom & Heymann), 392.
- Electron diffraction study of the structure of thin films of normal paraffins (Brummage), 414.
- Electron traps and dielectric changes in phosphorescent solids (Garlick & Gibson), 485.
- Energy distribution and stability of electrons in electric fields (Fröhlich), 532.
- von Engel, A. & Wills, M. S. Transverse magnetization in ferromagnetic crystals in relation to domain structure, 464.
- Ethylene, mechanism of catalytic oxidation (Twigg), 92, 105, 123.
- Evans, U. R. & Simnad, M. Tchorabdj. The mechanism of corrosion fatigue of mild steel, 372.
- Explosives, liquid, detonation (Bowden *et al.*), 291, 329.
- Fremberg, N. E. Some application of the Riesz potential to the theory of the electromagnetic field and the meson field, 18.
- Fröhlich, H. Energy distribution and stability of electrons in electric fields, 532.
- Fröhlich, H. On the theory of dielectric breakdown in solids, 521.
- Garlick, G. F. J. & Gibson, A. F. Electron traps and dielectric changes in phosphorescent solids, 485.
- General kinetic theory of liquids. I. The molecular distribution functions (Born & Green), 10.
- Gibson, A. F. *See* Garlick & Gibson.
- Green, H. S. *See* Born & Green.
- Heymann, E. *See* Bloom & Heymann.
- Hill, R., Lee, E. H. & Tupper, S. J. The theory of wedge indentation of ductile materials, 273.
- Hinshelwood, C. N. The more recent work on the reaction between hydrogen and oxygen, Bakerian Lecture, 1.
- Honeycombe, R. W. K. *See* Boas & Honeycombe.
- Hot spots on rubbing surfaces and the detonation of explosives by friction (Bowden, Stone & Tudor), 329.
- Hydrogen and oxygen, reaction (Hinshelwood), 1.
- Intensities of γ rays, studied by means of their Compton secondaries (Siegbahn), 541.
- Internal ballistics of a gun after shot ejection (Corner), 255.
- Internal ballistics of a leaking gun (Corner), 237.
- Ionic migration of molten salts (Bloom & Heymann), 392.
- Jeffrey, G. A. The structure of polyisoprenes. VI. An investigation of the molecular structure of dibenzyl by X-ray analysis, 222.
- Kinetic theory of liquids (Born & Green), 10.
- Lee, E. H. *See* Hill, Lee & Tupper.

- Measurement of the ultimate pressures of oil-diffusion pumps (Blears), 62.
Mechanism of corrosion fatigue of mild steel (Evans & Simnad), 372.
Mechanism of the catalytic oxidation of ethylene. I. Experiments using a flow system (Twigg), 92.
Mechanism of the catalytic oxidation of ethylene. II. Reactions between ethylene, etc. and chemisorbed oxygen monolayers (Twigg), 105.
Mechanism of the catalytic oxidation of ethylene. III. The reaction between ethylene and oxygen in a static system (Twigg), 123.
Meta-dinitrobenzene, crystal structure (Aroher), 51.
Morris, R. M. The two-dimensional hydrodynamical theory of moving aerofoils, 439.
Mulcahy, M. F. R. *See* Bowden, Mulcahy, Vines & Yoffe.
- Oil-diffusion pumps, measurement of ultimate pressures of (Blears), 62.
Owen, E. A. & Williams, R. W. The effect of temperature on the intensity of X-ray reflexion, 509.
- Period of impact, the time of initiation and the rate of growth of the explosion of nitroglycerine (Bowden, Mulcahy, Vines & Yoffe), 311.
Phosphorescent solids, electron traps and dielectric changes (Garlick & Gibson), 485.
Polyisoprenes, structure (Jeffrey), 222.
- Quantal calculation of the photo-ionization cross-section of atomic potassium (Bates), 350.
- Raman effect in crystals (Born & Bradburn), 161.
Reaction between hydrogen and oxygen, the more recent work on, Bakerian Lecture (Hinshelwood), 1.
Robinson, Sir Robert, Anniversary Address, 143.
- Saunders, D. H. The crystal structure of some molecular complexes of 4:4'-dinitrodiphenyl. I. The complex with 4-hydroxydiphenyl, 31.
Sheppard, P. A. The aerodynamic drag of the earth's surface and the value of von Karman's constant in the lower atmosphere, 208.
Siegbahn, K. Intensities of γ rays studied by means of their Compton secondaries, 541.
Simnad, M. Tchouabdjji. *See* Evans & Simnad.
Some applications of the Riesz potential to the theory of the electromagnetic field and the meson field (Fremberg), 18.
Stone, M. A. *See* Bowden, Stone & Tudor.
Structure and thermal properties associated with some hydrogen bonds in crystals. VII. Behaviour of KH_2PO_4 and KH_2AsO_4 on cooling (Ubbelohde & Woodward), 358.
Structure of polyisoprenes. VI. An investigation of the molecular structure of dibenzyl by X-ray analysis (Jeffrey), 222.
- Tchouabdjji Simnad, M. *See* Evans & Simnad.
Theory of dielectric breakdown in solids (Fröhlich), 521.
Theory of the Raman effect in crystals in particular rock-salt (Born & Bradburn), 161.
Theory of wedge indentation of ductile materials (Hill, Lee & Tupper), 273.
Thermal scattering of X-rays by crystals. I. Dynamical foundation (Begbie & Born), 179.
Thermal scattering of X-rays by crystals. II. The thermal scattering of the face-centred cubic and the close-packed hexagonal lattices (Begbie), 189.

Index

- Transverse magnetization in ferromagnetic crystals in relation to domain structure (von Engel & Wills), 464.
- Tudor, G. K. *See* Bowden, Stone & Tudor.
- Tupper, S. J. *See* Hill, Lee & Tupper.
- Twigg, G. H. The mechanism of the catalytic oxidation of ethylene. I. Experiments using a flow system, 92.
- Twigg, G. H. The mechanism of the catalytic oxidation of ethylene. II. Reactions between ethylene, etc. and chemisorbed oxygen monolayers, 105.
- Twigg, G. H. The mechanism of the catalytic oxidation of ethylene. III. The reactions between ethylene and oxygen in a static system, 123.
- Two-dimensional hydrodynamical theory of moving aerofoils (Morris), 439.
- Ubbelohde, A. R. & Woodward, I. Structure and thermal properties associated with some hydrogen bonds in crystals. VII. Behaviour of KH_2PO_4 and KH_2AsO_4 on cooling, 358.
- Vines, R. G. *See* Bowden, Mulcahy, Vines & Yoffe.
- Wedge indentation of ductile materials, theory (Hill *et al.*), 273.
- Williams, R. W. *See* Owen & Williams.
- Wills, M. S. *See* von Engel & Wills.
- Woodward, I. *See* Ubbelohde & Woodward.
- X-ray reflexion, effect of temperature (Owen & Williams), 509.
- X-rays, thermal scattering by crystals (Begbie), 189; (Begbie & Born), 179.
- Yoffe, A. *See* Bowden, Mulcahy, Vines & Yoffe.

L.A. 44-1. 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY, NEW DELHI.

[illegible]

GIPNLK—H-40 I.A.R.I.—29-4-5—15,000